

New York State Testing Program 2014: English Language Arts and Mathematics Grades 3–8



Technical Report

**Pearson
2014**

Developed and published under contract with the New York State Education Department by Pearson, 2510 North Dodge Street, Iowa City, Iowa 52245. Copyright © 2014 by the New York State Education Department.

Permission is hereby granted for New York State School administrators and educators to reproduce these materials, located online at <http://www.p12.nysed.gov/assessment/reports/>, in the quantities necessary for their school's use, but not for sale, provided copyright notices are retained as they appear in these publications. This permission does not apply to distribution of these materials, electronically or by any other means, other than for school use.

Table of Contents

Section 1: Introduction and Overview	1
Introduction.....	1
Test Purpose	1
Target Population	1
Test Use and Decisions Based on Assessment	1
Scale Scores	2
Statewide Percentile Ranks	2
Proficiency Level Cut Scores and Classification	2
Subscores	3
Testing Accommodations	4
Test Transcriptions.....	5
Test Translations.....	5
Section 2: Test Design and Development	6
Test Descriptions.....	6
ELA Tests	6
Mathematics Tests	7
Test Configuration.....	7
Test Book Design and Testing Times	7
Embedded Field Test Items	8
New York State Educators’ Involvement in Test Development	8
Development and Review Process	9
Test Blueprints	9
Passage Selection and Item Criteria Documents	9
Passage Finding.....	12
Item Development	12
Educator Item Review	13
Field Testing	14
Rangefinding	15
Item Selection and Test Creation (Criteria and Process)	15
Test Form Production and Reviews.....	16
Final Eyes Committee.....	16
Proficiency and Performance Standards.....	16
Section 3: Validity	17
Content Validity.....	17
Construct (Internal Structure) Validity	18
Internal Consistency.....	18
Unidimensionality.....	18
Detection of Bias.....	22

Section 4: Test Administration and Scoring	24
Test Administration	24
Scoring Procedures of Operational Tests	24
Scoring Models	25
Scoring of Constructed-Response Items	25
Scorer Qualifications and Training	26
Quality Control Process	26
Section 5: Operational Test Data Collection and Classical Analysis	27
Data Collection	27
Data Processing	27
Classical Analysis and Calibration Sample Characteristics	33
Classical Data Analysis	44
Item Difficulty and Point Biserial Correlation Coefficients	45
Speededness	46
Differential Item Functioning	48
Section 6: IRT Calibration and Equating	51
IRT Models and Rationale for Use	51
Calibration Sample	52
Calibration Process	58
Item-Model Fit	60
Local Independence	61
Equating and Scaling	62
Anchor Set Evaluation	65
Test Characteristic Curves	67
Scoring Procedure	80
Raw Score-to-Scale Score and SEM Conversion Tables	80
Section 7: Reliability and Standard Error of Measurement	83
Test Reliability	83
Test Statistics and Reliability for Total Test	83
Reliability of MC Items	85
Reliability of CR Items	85
Test Reliability for Reporting Categories	86
Standard Error of Measurement	97
Performance Level Classification Consistency and Accuracy	97
Consistency	98
Accuracy	99
Section 8: Summary of Operational Test Results	101
Scale Score Distribution Summary	101
ELA Scale Score and Subscore Distributions	101

ELA Grade 3	103
ELA Grade 4	104
ELA Grade 5	105
ELA Grade 6	106
ELA Grade 7	107
ELA Grade 8	108
Mathematics Scale Score Distributions	109
Mathematics Grade 3	110
Mathematics Grade 4	111
Mathematics Grade 5	113
Mathematics Grade 6	114
Mathematics Grade 7	115
Mathematics Grade 8	116
Performance Level Distribution Summary	118
ELA Test Performance Level Distributions	118
ELA Grade 3	119
ELA Grade 4	120
ELA Grade 5	121
ELA Grade 6	122
ELA Grade 7	123
ELA Grade 8	124
Mathematics Test Performance Level Distributions	125
Mathematics Grade 3	125
Mathematics Grade 4	126
Mathematics Grade 5	128
Mathematics Grade 6	129
Mathematics Grade 7	130
Mathematics Grade 8	131
References	133
Appendix A: ELA and Mathematics Test Configuration and Testing Times	136
Appendix B: ELA and Mathematics Test Blueprints	141
Appendix C: Passage Selection Guidelines for Assessing ELA	143
Appendix D: Universal Design Item Checklist	144
Appendix E: Criteria for Item Acceptability	147
Appendix F: Psychometric Guidelines for Operational Item Selection	149
Appendix G: Operational Item Maps	150
Appendix H: ELA Short-Response Rubric	167

Appendix I: ELA Extended-Response Rubric	168
Appendix J: Mathematics Short-Response Rubric	171
Appendix K: Mathematics Extended-Response Rubric.....	172
Appendix L: Factor Analysis Results for Select Subgroups	173
Appendix M: Classical Test Theory Statistics.....	189
Appendix N: Items Flagged for DIF.....	206
Appendix O: Item Response Theory Statistics.....	208
Appendix P: Derivation and Estimation of Classification Consistency and Accuracy	241
Appendix Q: Raw Score-to-Scale Score and Scale Score Frequency Tables	244

List of Tables

Table 1. ELA Subscore Categories and Total Possible Score Points	4
Table 2. Mathematics Subscore Categories and Total Possible Score Points	4
Table 3. Summary of Unique 2013 Field Test Items.....	14
Table 4. ELA Tests Factor Analysis	20
Table 5. Mathematics Tests Factor Analysis	21
Table 6A. ELA Grade 3 Data Cleaning.....	29
Table 6B. ELA Grade 4 Data Cleaning	29
Table 6C. ELA Grade 5 Data Cleaning	29
Table 6D. ELA Grade 6 Data Cleaning.....	30
Table 6E. ELA Grade 7 Data Cleaning	30
Table 6F. ELA Grade 8 Data Cleaning.....	30
Table 7A. Mathematics Grade 3 Data Cleaning	31
Table 7B. Mathematics Grade 4 Data Cleaning	31
Table 7C. Mathematics Grade 5 Data Cleaning	31
Table 7D. Mathematics Grade 6 Data Cleaning	32
Table 7E. Mathematics Grade 7 Data Cleaning.....	32
Table 7F. Mathematics Grade 8 Data Cleaning.....	32
Table 8A. ELA Grade 3 Sample Characteristics	33
Table 8B. ELA Grade 4 Sample Characteristics	34
Table 8C. ELA Grade 5 Sample Characteristics	35
Table 8D. ELA Grade 6 Sample Characteristics	36
Table 8E. ELA Grade 7 Sample Characteristics.....	37
Table 8F. ELA Grade 8 Sample Characteristics.....	38
Table 9A. Mathematics Grade 3 Sample Characteristics	39
Table 9B. Mathematics Grade 4 Sample Characteristics.....	40
Table 9C. Mathematics Grade 5 Sample Characteristics.....	41
Table 9D. Mathematics Grade 6 Sample Characteristics	42
Table 9E. Mathematics Grade 7 Sample Characteristics.....	43
Table 9F. Mathematics Grade 8 Sample Characteristics.....	44
Table 10A. ELA Correlations (across Books)	47
Table 10B. Mathematics Correlations (across Books)	47
Table 11A. ELA Correlations (Book by Book).....	48
Table 11B. Mathematics Correlations (Book by Book)	48
Table 12. ELA Classical DIF Sample N-Counts	49
Table 13. Mathematics Classical DIF Sample N-Counts	50
Table 14. ELA Items Flagged for DIF.....	50
Table 15. Mathematics Items Flagged for DIF.....	50
Table 16A. ELA Grades 3 and 4 Demographic Statistics	53

Table 16B. ELA Grades 5 and 6 Demographic Statistics.....	54
Table 16C. ELA Grades 7 and 8 Demographic Statistics.....	55
Table 17A. Mathematics Grades 3 and 4 Demographic Statistics.....	56
Table 17B. Mathematics Grades 5 and 6 Demographic Statistics.....	57
Table 17C. Mathematics Grades 7 and 8 Demographic Statistics.....	58
Table 18. ELA Calibration Results.....	59
Table 19. Mathematics Calibration Results	59
Table 20. ELA Equating Coefficients.....	63
Table 21. Mathematics Equating Coefficients.....	63
Table 22. ELA Scaling Coefficients	64
Table 23. Mathematics Scaling Coefficients	65
Table 24. ELA Performance-Level Cut Scores	81
Table 25. Mathematics Performance-Level Cut Scores	82
Table 26. ELA Test Form Statistics.....	84
Table 27. ELA Test Reliability and Standard Error of Measurement	84
Table 28. Mathematics Test Form Statistics.....	84
Table 29. Mathematics Test Reliability and Standard Error of Measurement	84
Table 30. ELA MC Item Reliability and Standard Error of Measurement.....	85
Table 31. Mathematics MC Item Reliability and Standard Error of Measurement.....	85
Table 32. ELA CR Item Reliability and Standard Error of Measurement.....	85
Table 33. Mathematics CR Item Reliability and Standard Error of Measurement.....	86
Table 34A. ELA Grade 3 Test Reliability by Subgroup.....	87
Table 34B. ELA Grade 4 Test Reliability by Subgroup.....	87
Table 34E. ELA Grade 7 Test Reliability by Subgroup	89
Table 34F. ELA Grade 8 Test Reliability by Subgroup	90
Table 35A. Mathematics Grade 3 Test Reliability by Subgroup.....	91
Table 35B. Mathematics Grade 4 Test Reliability by Subgroup	92
Table 35C. Mathematics Grade 5 Test Reliability by Subgroup	93
Table 35D. Mathematics Grade 6 Test Reliability by Subgroup.....	94
Table 35E. Mathematics Grade 7 Test Reliability by Subgroup	95
Table 35F. Mathematics Grade 8 Test Reliability by Subgroup	96
Table 36. Decision Consistency (All Cuts).....	98
Table 37. Decision Consistency (Level III Cut)	99
Table 38. Decision Agreement (Accuracy) Estimates	100
Table 39. ELA Scale Score Distribution Summary	101
Table 40. ELA Subscore Summary	102
Table 41A. ELA Grade 3 Scale Score Distribution by Subgroup	103
Table 41B. ELA Grade 4 Scale Score Distribution by Subgroup.....	104
Table 41C. ELA Grade 5 Scale Score Distribution by Subgroup.....	105
Table 41D. ELA Grade 6 Scale Score Distribution by Subgroup	106

Table 41E. ELA Grade 7 Scale Score Distribution by Subgroup	107
Table 41F. ELA Grade 8 Scale Score Distribution by Subgroup	108
Table 42. Mathematics Scale Score Distribution Summary	109
Table 43. Mathematics Subscore Summary.....	110
Table 44A. Mathematics Grade 3 Scale Score Distribution by Subgroup	111
Table 44B. Mathematics Grade 4 Scale Score Distribution by Subgroup.....	112
Table 44C. Mathematics Grade 5 Scale Score Distribution by Subgroup.....	113
Table 44D. Mathematics Grade 6 Scale Score Distribution by Subgroup	114
Table 44E. Mathematics Grade 7 Scale Score Distribution by Subgroup.....	116
Table 44F. Mathematics Grade 8 Scale Score Distribution by Subgroup	117
Table 45. ELA Test Performance Level Distributions	118
Table 46A. ELA Grade 3 Performance Level Distribution by Subgroup.....	119
Table 46B. ELA Grade 4 Performance Level Distribution by Subgroup.....	120
Table 46C. ELA Grade 5 Performance Level Distribution by Subgroup.....	121
Table 46D. ELA Grade 6 Performance Level Distribution by Subgroup.....	122
Table 46E. ELA Grade 7 Performance Level Distribution by Subgroup	123
Table 46F. ELA Grade 8 Performance Level Distribution by Subgroup	124
Table 47. Mathematics Test Performance Level Distributions	125
Table 48A. Mathematics Grade 3 Performance Level Distribution by Subgroup.....	126
Table 48B. Mathematics Grade 4 Performance Level Distribution by Subgroup.....	127
Table 48C. Mathematics Grade 5 Performance Level Distribution by Subgroup.....	128
Table 48D. Mathematics Grade 6 Performance Level Distribution by Subgroup.....	129
Table 48E. Mathematics Grade 7 Performance Level Distribution by Subgroup	131
Table 48F. Mathematics Grade 8 Performance Level Distribution by Subgroup	132
Table A1. ELA Test Configuration	136
Table A2. Mathematics Test Configuration.....	137
Table A3. ELA Testing Times.....	138
Table A4. Math Testing Times	139
Table B1. ELA Test Blueprint	141
Table B2. Mathematics Test Blueprint	142
Table G1. ELA Grade 3 Operational Item Map	150
Table G2. ELA Grade 4 Operational Item Map	151
Table G3. ELA Grade 5 Operational Item Map	152
Table G4. ELA Grade 6 Operational Item Map	153
Table G5. ELA Grade 7 Operational Item Map	155
Table G6. ELA Grade 8 Operational Item Map	156
Table G7. Mathematics Grade 3 Operational Item Map.....	157
Table G8. Mathematics Grade 4 Operational Item Map.....	159
Table G9. Mathematics Grade 5 Operational Item Map.....	160
Table G10. Mathematics Grade 6 Operational Item Map.....	162

Table G11. Mathematics Grade 7 Operational Item Map.....	163
Table G12. Mathematics Grade 8 Operational Item Map.....	165
Table L1. ELA Grade 3 Test Factor Analyses for Select Subgroups	173
Table L2. ELA Grade 4 Test Factor Analyses for Select Subgroups	174
Table L3. ELA Grade 5 Test Factor Analyses for Select Subgroups	176
Table L4. ELA Grade 6 Test Factor Analyses for Select Subgroups	177
Table L5. ELA Grade 7 Test Factor Analyses for Select Subgroups	179
Table L6. ELA Grade 8 Test Factor Analyses for Select Subgroups	180
Table L7. Mathematics Grade 3 Test Factor Analyses for Select Subgroups	182
Table L8. Mathematics Grade 4 Test Factor Analyses for Select Subgroups	183
Table L9. Mathematics Grade 5 Test Factor Analyses for Select Subgroups	184
Table L10. Mathematics Grade 6 Test Factor Analyses for Select Subgroups	185
Table L11. Mathematics Grade 7 Test Factor Analyses for Select Subgroups	186
Table L12. Mathematics Grade 8 Test Factor Analyses for Select Subgroups	187
Table M1. ELA Grade 3 Classical Item Analysis.....	189
Table M2. ELA Grade 4 Classical Item Analysis.....	190
Table M3. ELA Grade 5 Classical Item Analysis.....	191
Table M4. ELA Grade 6 Classical Item Analysis.....	193
Table M5. ELA Grade 7 Classical Item Analysis.....	194
Table M6. ELA Grade 8 Classical Item Analysis.....	195
Table M7. Mathematics Grade 3 Classical Item Analysis.....	197
Table M8. Mathematics Grade 4 Classical Item Analysis.....	198
Table M9. Mathematics Grade 5 Classical Item Analysis.....	199
Table M10. Mathematics Grade 6 Classical Item Analysis.....	201
Table M11. Mathematics Grade 7 Classical Item Analysis.....	202
Table M12. Mathematics Grade 8 Classical Item Analysis.....	204
Table N1. ELA MC Item Classical DIF Flags.....	206
Table N2. ELA CR Item Classical DIF Flags.....	207
Table N3. Mathematics MC Item Classical DIF Flags.....	207
Table N4. Mathematics CR Item Classical DIF Flags.....	207
Table O1. ELA Grade 3 Item Fit Statistics.....	208
Table O2. ELA Grade 4 Item Fit Statistics.....	209
Table O3. ELA Grade 5 Item Fit Statistics.....	210
Table O4. ELA Grade 6 Item Fit Statistics.....	212
Table O5. ELA Grade 7 Item Fit Statistics.....	213
Table O6. ELA Grade 8 Item Fit Statistics.....	214
Table O7. Mathematics Grade 3 Item Fit Statistics.....	216
Table O8. Mathematics Grade 4 Item Fit Statistics.....	217
Table O9. Mathematics Grade 5 Item Fit Statistics.....	218
Table O10. Mathematics Grade 6 Item Fit Statistics.....	220

Table O11. Mathematics Grade 7 Item Fit Statistics	221
Table O12. Mathematics Grade 8 Item Fit Statistics	223
Table O13. ELA Grade 3 Operational Item Parameter Estimates	224
Table O14. ELA Grade 4 Operational Item Parameter Estimates	225
Table O15. ELA Grade 5 Operational Item Parameter Estimates	226
Table O16. ELA Grade 6 Operational Item Parameter Estimates	228
Table O17. ELA Grade 7 Operational Item Parameter Estimates	229
Table O18. ELA Grade 8 Operational Item Parameter Estimates	230
Table O19. Mathematics Grade 3 Operational Item Parameter Estimates	232
Table O20. Mathematics Grade 4 Operational Item Parameter Estimates	233
Table O21. Mathematics Grade 5 Operational Item Parameter Estimates	234
Table O22. Mathematics Grade 6 Operational Item Parameter Estimates	236
Table O23. Mathematics Grade 7 Operational Item Parameter Estimates	237
Table O24. Mathematics Grade 8 Operational Item Parameter Estimates	239
Table Q1. ELA Grade 3 Raw Score-to-Scale Score Table	244
Table Q2. ELA Grade 4 Raw Score-to-Scale Score Table	245
Table Q3. ELA Grade 5 Raw Score-to-Scale Score Table	246
Table Q4. ELA Grade 6 Raw Score-to-Scale Score Table	247
Table Q5. ELA Grade 7 Raw Score-to-Scale Score Table	248
Table Q6. ELA Grade 8 Raw Score-to-Scale Score Table	249
Table Q7. Mathematics Grade 3 Raw Score-to-Scale Score Table	250
Table Q8. Mathematics Grade 4 Raw Score-to-Scale Score Table	251
Table Q9. Mathematics Grade 5 Raw Score-to-Scale Score Table	252
Table Q10. Mathematics Grade 6 Raw Score-to-Scale Score Table	253
Table Q11. Mathematics Grade 7 Raw Score-to-Scale Score Table	254
Table Q12. Mathematics Grade 8 Raw Score-to-Scale Score Table	255
Table Q13. ELA Grade 3 Scale Score Frequency Distribution	256
Table Q14. ELA Grade 4 Scale Score Frequency Distribution	257
Table Q15. ELA Grade 5 Scale Score Frequency Distribution	259
Table Q16. ELA Grade 6 Scale Score Frequency Distribution	261
Table Q17. ELA Grade 7 Scale Score Frequency Distribution	263
Table Q18. ELA Grade 8 Scale Score Frequency Distribution	265
Table Q19. Mathematics Grade 3 Scale Score Frequency Distribution	267
Table Q20. Mathematics Grade 4 Scale Score Frequency Distribution	269
Table Q21. Mathematics Grade 5 Scale Score Frequency Distribution	271
Table Q22. Mathematics Grade 6 Scale Score Frequency Distribution	273
Table Q23. Mathematics Grade 7 Scale Score Frequency Distribution	275
Table Q24. Mathematics Grade 8 Scale Score Frequency Distribution	277

List of Figures

Figure 1. ELA Grade 3 TCC.....	68
Figure 2. ELA Grade 3 CSEM Curve.....	68
Figure 3. ELA Grade 4 TCC.....	69
Figure 4. ELA Grade 4 CSEM Curve.....	69
Figure 5. ELA Grade 5 TCC.....	70
Figure 6. ELA Grade 5 CSEM Curve.....	70
Figure 7. ELA Grade 6 TCC.....	71
Figure 8. ELA Grade 6 CSEM Curve.....	71
Figure 9. ELA Grade 7 TCC.....	72
Figure 10. ELA Grade 7 CSEM Curve.....	72
Figure 11. ELA Grade 8 TCC.....	73
Figure 12. ELA Grade 8 CSEM Curve.....	73
Figure 13. Mathematics Grade 3 TCC.....	74
Figure 14. Mathematics Grade 3 CSEM Curve.....	74
Figure 15. Mathematics Grade 4 TCC.....	75
Figure 16. Mathematics Grade 4 CSEM Curve.....	75
Figure 17. Mathematics Grade 5 TCC.....	76
Figure 18. Mathematics Grade 5 CSEM Curve.....	76
Figure 19. Mathematics Grade 6 TCC.....	77
Figure 20. Mathematics Grade 6 CSEM Curve.....	77
Figure 21. Mathematics Grade 7 TCC.....	78
Figure 22. Mathematics Grade 7 CSEM Curve.....	78
Figure 23. Mathematics Grade 8 TCC.....	79
Figure 24. Mathematics Grade 8 CSEM Curve.....	79

Section 1: Introduction and Overview

Introduction

This technical report provides detailed information regarding the technical, statistical, and measurement attributes of the New York State Testing Program (NYSTP) for the Grades 3–8 Common Core English Language Arts (ELA) and Mathematics 2014 Operational Tests. This report includes information about test content and test development, item (i.e., individual test question) and test statistics, validity and reliability, differential item functioning studies, test administration, scoring, equating, scaling, and student performance.

Test Purpose

The 2014 Grades 3–8 Common Core ELA and Mathematics NYSTP has been designed to measure student knowledge and skills as defined by grade-level New York State Common Core Learning Standards (CCLS) in ELA and Mathematics. The tests are designed to allow the classification of student proficiency into four performance levels. Likewise, the test provides students at each of these performance levels opportunities to demonstrate their knowledge and skills in the CCLS. Details about the content standards for ELA and mathematics are described in Section 2, subsections “Development and Review Process,” and “Test Blueprints.”

Target Population

Students in New York State public school Grades 3, 4, 5, 6, 7, and 8 (and ungraded students of equivalent chronological ages) are the target population for the Grades 3–8 NYSTP. Non-public schools may participate in the testing program, but their participation is not mandatory. In 2014, some non-public schools participated in the testing program across all grade levels. These schools were included in the data analyses. Public school students were required to take all State assessments administered at their grade level, except for a very small percentage of students with severe cognitive disabilities who took the New York State Alternate Assessment (NYSAA) for students with severe disabilities. For more detail on this exemption, please refer to the *NYSTP Grades 3–8 Common Core English Language Arts and Mathematics Tests School Administrator’s Manual* (SAM), available online at:

<http://www.p12.nysed.gov/assessment/sam/ei/eisam14rev.pdf>.

Test Use and Decisions Based on Assessment

The NYSTP Grades 3–8 Common Core ELA and Mathematics Tests are used to measure the extent to which individual students achieve the New York State Common Core Learning Standards in ELA and Mathematics, respectively, in order to determine whether schools, districts, and the State meet the required progress targets specified in the New York State accountability system. There are several types of scores available from the Grades 3–8 ELA and Mathematics Tests, and they are discussed in this section.

Scale Scores

The scale scores are a quantification of the proficiency measured by the Grades 3–8 Common Core ELA and Mathematics Tests at each grade level. Scale scores are comparable only within a given subject and grade. Scale scores are not comparable across grades or across subjects. The scale scores are reported at the individual student level and can be aggregated. Detailed information on the derivation and properties of the scale scores is provided in Section 6, “IRT Calibration and Equating.” The Grades 3–8 ELA and Mathematics Tests’ scale scores are the basis for placing students into proficiency levels, which are used to determine student progress within schools and districts, support registration of schools and districts, determine eligibility of students for additional educational services, and provide teachers with indicators of a student’s need, or lack of need, for remediation in specific content-area knowledge.

Statewide Percentile Ranks

Students’ scale scores were also presented as percentile ranks in order to indicate students’ performance relative to the entire testing population on a scale that may be more familiar than the operational test’s scale. Such statistics were estimated based on the how often each student earned a given scale score and thus present similar information as the scale score itself, but on an alternate scale.

Proficiency Level Cut Scores and Classification

Students are classified as Level I, Level II, Level III, and Level IV for the Grades 3–8 Common Core ELA and Mathematics Tests. The definitions of performance levels for the Grades 3–8 Common Core ELA Tests are as follows:

NYS Level I: Students performing at this level are well below proficient in standards for their grade. They demonstrate limited knowledge, skills, and practices embodied by the New York State P–12 Common Core Learning Standards for English Language Arts/Literacy that are considered insufficient for the expectations at this grade.

NYS Level II: Students performing at this level are below proficient in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Common Core Learning Standards for English Language Arts/Literacy that are considered partial but insufficient for the expectations at this grade.

NYS Level III: Students performing at this level are proficient in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Common Core Learning Standards for English Language Arts/Literacy that are considered sufficient for the expectations at this grade.

NYS Level IV: Students performing at this level excel in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Common Core Learning Standards for English Language Arts/Literacy that are considered more than sufficient for the expectations at this grade.

The definitions of performance levels for the Grades 3–8 Common Core Mathematics Tests are as follows:

NYS Level I: Students performing at this level are well below proficient in standards for their grade. They demonstrate limited knowledge, skills, and practices embodied by the New York State P–12 Common Core Learning Standards for Mathematics that are considered insufficient for the expectations at this grade.

NYS Level II: Students performing at this level are below proficient in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Common Core Learning Standards for Mathematics that are considered partial, but insufficient for the expectations at this grade.

NYS Level III: Students performing at this level are proficient in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Common Core Learning Standards for Mathematics that are considered sufficient for the expectations at this grade.

NYS Level IV: Students performing at this level excel in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Common Core Learning Standards for Mathematics that are considered more than sufficient for the expectations at this grade.

The performance level cut scores used to distinguish between Levels I, II, III, and IV were established during the process of standard-setting in Summer 2013. The process is described in detail in Section 8 and Appendix P in the 2013 technical report (NYSED, 2014).

Subscores

The Grades 3–8 Common Core ELA tests have two subscores: reading (which includes all multiple-choice items assessing both reading and language standards); and writing to sources (which includes all constructed-response items assessing reading, writing, and language standards). The Grades 3–8 Common Core mathematics tests have three subscores. The mathematics subscores are the domain level scores for questions measuring the *Major Clusters* in each grade. The CCLS are divided into *Major*, *Supporting*, and *Additional Clusters*. Standards within *Major Clusters* are the intended focus of instruction and assessment and account for the majority of the mathematics test items. The *Supporting* and *Additional Clusters* are mathematics standards that serve to both introduce and reinforce *Major Clusters*. Table 1 and Table 2 below present the reporting subscore categories and the point values that correspond to each on the 2014 tests. In 2014, subscores were reported in two ways: (1) a raw score (i.e., number of points earned) out of the total score on the test; and (2) the average score at the state level for each subscore category.

Table 1. ELA Subscore Categories and Total Possible Score Points

Grade	Total Subscore Points	
	Reading	Writing to Sources
3	29	20
4	31	24
5	42	24
6	42	24
7	41	24
8	42	24

Table 2. Mathematics Subscore Categories and Total Possible Score Points

Grade	Reporting Subscores		
	Subscore 1	Subscore 2	Subscore 3
3	Operations and Algebraic Thinking 27	Number and Operations–Fractions 12	Measurement and Data 12
4	Operations and Algebraic Thinking 11	Number and Operations in Base Ten 17	Number and Operations–Fractions 18
5	Number and Operations in Base Ten 18	Number and Operations–Fractions 25	Measurement and Data 10
6	Ratios and Proportional Relationships 18	The Number System 12	Expressions and Equations 28
7	Ratios and Proportional Relationships 20	The Number System 14	Expressions and Equations 22
8	Expressions and Equations 30	Functions 19	Geometry 12

Testing Accommodations

In accordance with federal law under the Americans with Disabilities Act and the section, Fairness in Testing, as outlined by the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014), accommodations that do not alter the measurement of any construct being tested are allowed for test takers. The allowance is in accordance with a student's Individualized Education Program (IEP) or Section 504 Accommodation Plan (504 Plan). School principals are responsible for ensuring that proper accommodations are provided when necessary and that staff providing accommodations are

properly trained. Details on testing accommodations can be found in the 2014 School Administrator's Manual (SAM).

Test Transcriptions

For visually impaired students, large-type and Braille editions of the test books are provided. In most cases, the students dictate and/or record their responses, the teachers transcribe student responses to the multiple-choice (MC) items onto scannable answer sheets, and the teachers transcribe the responses to the constructed-response (CR) items onto the regular test books. Some of the students who use large-type editions will fill in the answer sheets by themselves. The large-type editions are created by Pearson and printed by the New York State Education Department (NYSED), and the Braille editions are produced by gh, LLC. gh employs certified Library of Congress Braille transcribers and delivers Braille in accordance to the Braille Authority of North America (BANA) standard. Camera-copy versions of the regular test books are provided to the Braille vendor, which then produces the Braille editions. Proofs of the Braille editions are submitted to NYSED for review and approval prior to production.

Test Translations

The NYSTP Grades 3–8 Common Core Mathematics Tests are translated into five languages: Chinese, Haitian-Creole, Korean, Russian, and Spanish. These tests are translated to provide students the opportunity to demonstrate mathematical proficiency independent of their command of the English language. Sample tests are available in each translated language at the following location:

<http://www.p12.nysed.gov/assessment/math/samplers/>

English language learners taking the Grades 3–8 Common Core Mathematics Tests may be provided with an oral translation of the test when a written translation is not available in the student's native language. Additionally, the following testing accommodations were made available to English language learners: time extension, separate testing location, bilingual glossaries, simultaneous use of English and alternative-language editions, oral translation for lower-incidence languages, and writing responses in the native language.

The NYSTP Grades 3–8 Common Core ELA Tests are not translated into any other language because they are assessments of proficiency in English language arts.

Section 2: Test Design and Development

Test Descriptions

The 2014 Grades 3–8 Common Core ELA and Mathematics Tests are criterion-referenced tests composed of multiple-choice (MC) and constructed-response (CR) test items based on the New York State P–12 Common Core Learning Standards (CCLS). The tests were administered in New York State classrooms during April 2014 over a three-day period. Details on the administration and scoring of these tests can be found in Section 4, “Test Administration and Scoring.” Additional information can be found in the *NYSTP Grades 3–8 Common Core English Language Arts and Mathematics Tests School Administrator’s Manual* (SAM), available at:

<http://www.p12.nysed.gov/assessment/sam/ei/eisam14rev.pdf>.

ELA Tests

The 2014 Grade 3–8 Common Core English Language Arts Tests were designed to measure student literacy as defined by the CCLS. The tests assessed Reading, Writing, and Language standards using multiple-choice, short-response, and extended-response questions. All questions were based on close reading of informational, literary, or paired texts. All texts were drawn from authentic, grade-level works.

Multiple-choice questions were designed to assess Common Core Reading and Language Standards. Multiple-choice questions require students to analyze different aspects of a given text, including central idea, style elements, character and plot development, and vocabulary.

Short-response questions were designed to assess Common Core Reading and Language Standards. These were single questions in which students use textual evidence to support their answer to an inferential question. These questions asked students to make an inference, state a position, or draw a conclusion based on their analysis of the passage and then provide two pieces of text-based evidence to support their answers. In responding to these questions, students were expected to write in complete sentences. The rubric for the short-response items can be found in Appendix H.

Extended-response questions were designed to assess Reading, Writing, and Language Standards, with a focus primarily on the Writing Standard. Extended-response questions required comprehension and analysis of either an individual text or paired texts. Paired texts required students to read and analyze two related texts. Paired texts were related by theme, genre, tone, time period, or other characteristics. Many extended-response questions asked students to express a position and support it with text-based evidence. For paired texts, students were expected to synthesize ideas between and draw evidence from both texts. Extended-response questions required students to demonstrate their ability to write a coherent essay using textual evidence to support their ideas. The rubric for the extended-response items can be found in Appendix I.

Mathematics Tests

The 2014 Grade 3–8 Common Core Mathematics Tests were designed to measure student mathematic understanding as defined by the CCLS. The tests required that students understand mathematics conceptually, use prerequisite skills with grade-level mathematical facts, decide which formulas and tools (e.g., protractors and rulers) to use, and solve mathematic problems rooted in the real world. The tests contained multiple-choice, short-response (2-point), and extended-response (3-point) questions. For multiple-choice questions, students selected the correct response from four answer choices. For short- and extended-response questions, students wrote an answer to an open-ended question. Some questions required students to show their work or to explain, in words, how they arrived at their answers.

Mathematics multiple-choice questions were mainly used to assess standard algorithms and conceptual standards. Multiple-choice questions incorporated the New York State CCLS, some in real-world applications. Many multiple-choice questions required students to complete multiple steps. Likewise, many of these questions were linked to more than one standard, drawing on the simultaneous application of multiple skills and concepts.

Short-response questions were mainly used to assess conceptual and application standards. The questions required students to complete a task and show their work. Like multiple-choice questions, short-response questions often required multiple steps, the application of multiple mathematics skills, and real-world applications. The rubric for the mathematics short-response items can be found in Appendix J.

Extended-response questions were mainly used to assess students’ abilities to show their understanding of mathematical procedures, conceptual understanding, and application of those procedures and concepts. Extend-response questions required students to complete two or more tasks or a more extensive problem and show their work. Some questions also assessed student reasoning and the ability to critique the arguments of others. The rubric for the mathematics extended-response items can found in Appendix K.

Test Configuration

Test Book Design and Testing Times

The 2014 Grades 3–8 Common Core ELA Tests were composed of three books per grade and administered in three sessions over three days. Each day consisted of one book; Book 1 and Book 2 contained literary and informational reading passages and MC items based on the passages. Book 2 also contained reading passages with short-response items and an extended-response item based on those passages. Book 3 contained only reading passages with short-response items and an extended-response item based on those passages.

The 2014 Grades 3–8 Common Core Mathematics Tests were composed of three books per grade and administered in three sessions over three days. Each day consisted of one book. Book 1 and Book 2 contained MC items. Book 3 contained short- and extended-response items. The tables in Appendix A provide information on the numbers and types of items in each book for the Grades 3–8 Common Core ELA and Mathematics Tests and the testing times.

Embedded Field Test Items

In 2010, the Department announced its commitment to embed multiple-choice items for field-testing within the Spring 2012 Grades 3–8 ELA and Mathematics Operational Tests; this commitment continued for the Spring 2014 administrations of the Common Core assessments. Embedding field-test items allows for a better representation of student responses and provides more reliable field-test data on which to build future operational tests. In other words, since the specific locations of the embedded field-test items were not disclosed and they look the same as operational items, students were unable to differentiate field-test items from operational test items. Therefore, field-test data derived from embedded items are free of the effects of differential student motivation that may characterize stand-alone field-test designs. Embedding field-test items also reduced the number of stand-alone field-tests during the Spring of 2014 but did not eliminate the need for them.

New York State Educators' Involvement in Test Development

New York State educators are actively involved in Common Core ELA and Mathematics test development. New York State educators provide critical input throughout all stages of the test development process, which include Educator Item Review, Rangefinding, Final Eyes Meeting (a final review of the test books prior to printing) and Standard-Settings.

NYSED gathers a diverse group of educators to review all test materials in order to create fair and valid tests. The participants are selected for each testing activity based on:

- Certification and appropriate grade-level experience;
- Special population experience;
- Geographical region;
- Gender;
- Ethnicity;
- Type of school (urban, suburban, or rural).

The selected participants must be certified and have both teaching and testing experience. The majority of the participants are classroom teachers. In addition, specialists, such as reading coaches, literacy coaches, and special education and bilingual instructors participate as well. Some participants are also recommended by principals, professional organizations, Big Five Cities, the Staff and Curriculum Development Network (SCDN). A file of participants is maintained and is routinely updated with current participant information and the addition of possible future participants, as recruitment forms are received. This gives many educators the opportunity to participate in the test development process. Every effort is made to have diverse groups of educators participate in each testing event.

Additionally, Content Advisory Panels (CAPs) which are content-area-specific advisory panels made up of between 15 and 20 New York State P-20 educators whose members are nominated by state professional organizations, institutes of higher education, and educator unions, meet quarterly to review, vet, and provide comment on curricular and assessment work.

Development and Review Process

During the process of transitioning from legacy assessments to the new CCSS-aligned assessments, NYSED and Pearson sought consultation with the following:

- New York State Educators
- Student Achievement Partners
- College Board
- HumRRO (Human Resources Research Organization)

Test Blueprints

After careful consideration of administration constraints (i.e., feasibility of paper-based tests versus online tests, number and length of test forms, and location of multiple-choice and constructed-response items within test books) and timing constraints, the representation and distribution of content was determined.

The CCLS for ELA are organized into four strands: Reading, Writing, Language, and Speaking/Listening. Due to administration constraints, Speaking/Listening was determined to best be assessed in the classroom only; therefore, the Common Core ELA Tests assess three of the four strands: Reading, Writing, and Language. Content experts reviewed the Reading, Writing, and Language standards and recommended content coverage by standard and item-type based on the depth and breadth of each standard.

The CCLS for mathematics are divided into *standards*, *clusters*, and *domains*. *Standards* define what students should understand and be able to do and are further articulated into lettered *components*. *Clusters* are groups of related *standards*. *Domains* are larger groups of related *clusters* and *standards*. Content experts reviewed the mathematics standards and recommended content coverage by standard and item-type based on the emphasis of the cluster (*major*, *supporting*, *additional*) and depth and breadth of each standard.

Tables B1 and B2 in Appendix B show the test blueprint and actual number of score points in the Grades 3–8 Common Core ELA and Mathematics Tests, respectively. Included in the tables are the ranges of allowable points for each ELA Strand and mathematics Domain and the actual number of points on the 2014 operational tests.

Passage Selection and Item Criteria Documents

To guide test item development and to help ensure that NYS tests were measuring the CCLS for ELA and mathematics with fidelity, NYSED and Pearson established criteria for selecting passages and writing test items based on the consultation with the groups listed above.

The *Passage Selection Guidelines for Assessing CCSS ELA* were created to provide a framework that allows for the consistent selection of passages that are appropriately complex for the given grade; and contain the specific characteristics necessary to measure different standards (see Appendix C). The guidelines describe the quantitative methods used to determine the grade appropriateness of a given text. They also describe the grade specific text characteristics needed

to develop questions that measure any particular reading standard. The complete guidelines can be found here:

http://www.engageny.org/sites/default/files/resource/attachments/passage_selection_guidelines_for_assessing_ccss_ela.pdf.

Passage Review Criteria documents were created based on the passage selection guidelines and were used to evaluate each potential passage and determine whether it could be used to measure the CCSS for ELA. The criteria documents were used to determine whether each passage suggested for testing use was grade appropriate, fair, and possessed the necessary characteristics to assess each standard. Specifically, passages were evaluated for the presence and quality of key ideas and details, craft and structure, and integration of knowledge and ideas. The full passage review criteria can found here:

http://www.engageny.org/sites/default/files/resource/attachments/new_york_state_passage_review_criteria_protocol_document.doc.

Item Review Criteria for Grade 3–8 English Language Arts Tests were used to help ensure that each item was clear; was fair; measured a specific Common Core standard (or standards) with fidelity; and conformed to the specifications for the item type. Each section of the criteria includes pertinent questions used to determine whether or not an item was of sufficient quality so that it could move forward in the development process. The first two sections of the Item Review Criteria, Clarity and Fairness, identify the basic components of quality questions. The criteria for Clarity are used to help ensure that students understand what is asked in each question and that the language choice in the question does not negatively impact a student’s ability to perform the required task. For example, the criteria include checking to make sure the vocabulary of test items are on grade level and that questions avoid technical terms unrelated to the content. Likewise, the Fairness criteria are used to ensure questions are un-biased, non-offensive, and not disadvantageous to any given subgroup. The criteria also address how each item measures a given standard or standards, and articulates the aspects of each standard that the items need to address. Finally, the criteria establish key requirements for each item type, requiring, for example, that each two-point constructed-response question asks students to make a clear statement that can be supported with two independent text-based pieces of evidence. The complete English Language Arts Criteria documents can found here:

http://www.engageny.org/sites/default/files/resource/attachments/ela_item_review_criteria_grades_3-5.doc

and here:

http://www.engageny.org/sites/default/files/resource/attachments/ela_item_review_criteria_grades_6-8.doc.

Item Review Criteria for Grade 3–8 Mathematics Tests were used to ensure clarity, language and graphical appropriateness, fairness, freedom from bias, fidelity of measurement to CCSS, and conformity to the expectations for specific item types and formats for each test question. Each

section of the criteria includes pertinent questions that determine whether or not an item is of sufficient quality. The first two categories, Clarity and Graphical Appropriateness and Fairness, identify the basic components of quality assessment items. The criteria for Clarity and Graphical Appropriateness are used to help ensure that students understand what is asked in each question and that the language in the question does not adversely affect a student's ability to perform the required task. For example, the criteria include checking to make sure that the visual load for questions containing art are reasonable and that interpreting a graphic does not confuse the underlying construct. Likewise, the Fairness criteria are used to evaluate whether questions are un-biased, non-offensive, and not disadvantageous to any given subgroup(s). The criteria also require documentation of how each item measures the assigned mathematics standard(s). Finally, the criterion addresses the specific demands for different item types and formats. For example, the criteria for a three-point constructed-response item include making sure that items involve a multi-step processes and require students to show work. The complete math criteria can be found here:

http://www.engageny.org/sites/default/files/resource/attachments/math_item_review_criteria.doc.

The Multiple Representations for NYS Grade 3–8 Common Core Mathematics Tests document was developed to ensure the tests measured the deep conceptual understanding that CCSS demands rather than focusing on predictable math questions that require only algorithmic strategies to be solved correctly. Multiple Representations are a broad set of specifications that describe, refer, and symbolize the various, but not all, ways that math standards could be measured within the constraints of NYSTP. The document specifies three overarching families: procedural skills, conceptual understanding, and application. It also includes information about how to identify standards that might be measured using a particular representation and identifies types of math skills (e.g. application of process; explanation of a principle, etc.) that are appropriate to assess different representations. The full document can be found here:

http://www.engageny.org/sites/default/files/resource/attachments/mathematics_multiple_representations.pdf.

To create tests that were as equitable as possible for students, principles of universal design were employed during the creation of the tests and test questions. In a report published by the National Council on Educational Outcomes, “Universally designed assessments” are designed and developed from the beginning to allow participation of the widest possible range of students, and to result in valid inferences about performance for all students who participate in the assessment” (Thompson, S.J., Johnstone, C.J., & Thurlow, M.L. 2002). The report goes on to describe seven elements of a universally designed assessment. These elements are:

1. Inclusive assessment population
2. Precisely defined constructs
3. Accessible, non-biased items
4. Amenable to accommodations
5. Simple, clear, and intuitive instructions and procedures

6. Maximum readability and comprehensibility
7. Maximum legibility

In accordance with these elements, the checklist (Universal Design Item Checklist) in Appendix D was developed for use during item development.

Passage Finding

The goal of passage finding is to obtain high-quality texts from which to generate CCSS-aligned test questions. To do so, Pearson recruited independent passage finders and trained them using passage selection resources including the passage selection criteria. Passage finders were given assignments based on the test blueprint requirements. Passage finders submitted passages along with completed criteria documents and source information to Pearson ELA content specialists who reviewed the passages against the agreed upon criteria. Passages that did not meet the criteria were rejected, and passages that did meet the criteria were moved forward in the process where the text from scanned copies of the original sources was entered into templates. Once in the templates, readability metrics were determined for each text, and it was then proofread by Pearson copyeditors, fact checked by Pearson research librarians, reviewed for subject-specific content issues by science and social studies content specialists, and reviewed for universal-design issues by specifically trained Pearson reviewers. After the passages went through these review steps, Pearson ELA content specialists posted the passages and completed criteria documents for NYSED's review and approval for moving forward in the process.

NYSED staff retrieved the passages and criteria documents and reviewed both the passages and criteria documents. If the NYSED staff determined that a passage did not meet the criteria, the passage was rejected, and the NYSED staff provided Pearson with an explanation for the reason for rejection.

In addition to the content reviews performed by both Pearson and NYSED, the passages were also reviewed by executives in both organizations. The executive review focused on bias and sensitivity issues that were particular to New York State. Passages that passed both content and executive reviews were moved forward for item development.

Item Development

The goal of item development is to develop a sufficient number of high-quality, CCSS-aligned items to populate the test forms. Using the criteria documents for both subjects and the multiple-perspective document for mathematics, Pearson content leads trained item writers. The item writers had teaching or assessment experience in the subject area for which they were writing items, experience in writing for large-scale, high-stakes assessments, and a bachelor's degree in either education and/or the subject area for which they were assigned. The item writers were given specific assignments based on the test blueprint. For ELA, the item writers were also provided with the completed passage criteria documents.

The item writers provided items and completed criteria documents to Pearson content specialists for review. Two content specialists reviewed each item and its corresponding criteria document, and any items that did not meet the criteria were sent back to the writers with specific feedback

for revision. Items that did not meet the criteria after an attempted revision were rejected and replaced by Pearson content specialists. After the Pearson content specialists were satisfied that all of the items met the criteria, the items were reviewed by Pearson copyeditors. The mathematics items were also reviewed by subject-specific content specialists in science and social studies and by research librarians. The Pearson ELA and mathematics content specialists evaluated the feedback from the different internal groups and edited the items accordingly. Then the items and criteria documents were posted for NYSED's review and approval for moving forward in the process.

NYSED content experts retrieved the items and criteria documents and reviewed both the items and criteria documents. If NYSED staff determined that an item did not meet the criteria, the item was rejected, and the NYSED staff provided Pearson with an explanation for the reason for rejection. Pearson then replaced the item and completed criteria documents which were resubmitted to NYSED. If NYSED staff determined that an item met the criteria but could be improved with editing, the staff member recorded notes for the edits. Those notes were reviewed at face-to-face meetings where Pearson content staff and NYSED staff reviewed and edited all of the items to meet the criteria. All passages and items accepted at that meeting were moved forward for the Educator Item Review.

Educator Item Review

After items were reviewed by NYSED staff, the items were presented to panels of New York State educators. Based on their expertise, educators were assigned to grade-level and subject specific groups where they reviewed the items. The reviews were facilitated by Pearson content specialists and were also attended by NYSED staff. For ELA, reviewers first read and discussed the passages. For mathematics and ELA the educators then used the following checklist to review each item.

1. Does the item align to the designated standard(s)?
 - The item measures the content standard(s) it was designed to measure.
2. Does the item meet quality standards?
 - The item is worded clearly.
 - The reading level of the item is grade appropriate.
 - The item has one correct answer.
 - The item has plausible, unambiguous distractors.
 - All of the distractors are mutually exclusive.
3. Is the item fair?
 - The item is free from bias on the basis of students' personal characteristics such as gender or ethnicity.

As the educators reviewed the items, they discussed their judgments about them and if the educators felt that an item did not align to the standards, meet quality standards, or was not fair, they made recommendations for editing the item. NYSED staff and Pearson content specialists later reviewed the recommendations and made the appropriate edits.

Field Testing

Once the items have been developed and thoroughly reviewed by a variety of stakeholders, they must then be field-tested. The process of field-testing items is a critically important step in the test development process, as it is only through the gathering of actual student response data that a variety of psychometric characteristics may be evaluated. In particular, a summary of the unique items that passed the scrutiny of NYSED and Pearson content specialists as well as New York State educators and were field tested is included in Table 3. Note that more items were field tested than were needed on the operational forms because that enabled tests to be constructed with the best possible characteristics from both a content and psychometric perspective.

Table 3. Summary of Unique 2013 Field Test Items

Grade	Unique ELA Items by Type		Unique Mathematics Items by Type	
	Multiple-Choice	Constructed-Response	Multiple-Choice	Constructed-Response
3	126	48	96	22
4	125	48	120	25
5	138	48	120	25
6	137	48	132	25
7	137	48	132	25
8	137	48	135	25

Note. All constructed-response items were field tested under stand-alone conditions, while multiple-choice items were administered under both embedded and stand-alone conditions.

The first set of field test items were administered in the Spring of 2013 as embedded field test items within the 2013 operational test forms. As was noted above, the use of embedded field test items both yielded more reliable field-test data and reduced—but did not eliminate—the need for multiple-choice stand-alone field testing. One additional round of field testing was administered separately from the 2013 operational forms (i.e., as stand-alone tests)—later in the Spring of 2013.

In order to better understand how 2013 field test items may perform on future operational forms, a variety of analyses were conducted. All of the field test data underwent a series of representativeness checks. Because only a small sample of schools participate for any given subject and grade for stand-alone field testing, it was necessary to ensure that the stand-alone field test samples were representative of the entire State population in terms of: student achievement on prior years' tests, student gender, student ethnicity, and school Needs/Resource Capacity Category (NRC). Finally a variety of psychometric analyses were conducted, including classical item analysis, inter-rater reliability for constructed-response items, differential item functioning (DIF), item response theory (IRT) item calibration, equating, scaling and fit evaluation. Many of these analyses are described at length below; however inter-rater reliability analyses were not possible for the operational test, as only a single rater scored each constructed-response.

Rangefinding

After constructed-response items have been field-tested, rangefinding occurs. The purpose of rangefinding is to have New York State teachers review student constructed-responses and arrive at consensus scores based on the standards established by NYSED and the scoring rubrics. The consensus scores become the basis for operational rating guides and scoring ancillaries. To arrive at consensus, committees of New York State teachers review, discuss, and rate student responses to the field test questions. This process is overseen by NYSED content experts and Pearson Scoring Directors. The first step in the rangefinding process was to have the teacher committees review rubrics and a NYSED-approved grounding guide set, previously used for the 2013 field-test rangefinding sessions, to familiarize teachers with the application of NYSED standards and rubrics. The grounding guide sets contain student responses that illustrate the full range of scores on the rubric. The grounding guide sets are composed of student responses that had previously gone through the rangefinding process and been approved by NYSED to guide the scoring of field test and operational student responses. Referencing the previously approved guide set papers during the rangefinding sessions ensures consistency in the application of NYSED standards and rubrics from year-to-year.

After the committee reviewed the preapproved anchor set, the committee members familiarized themselves with each item type by group scoring a small number of responses that were representative of each of the different score points. Upon completion of the group-scoring exercise, committee members independently scored other student responses. After the independent scoring was completed, the committee reviewed and discussed their results and determined consensus scores for the responses. The rangefinding results were then used to build training materials for Pearson scorers who scored the field-test responses to constructed-response items.

Item Selection and Test Creation (Criteria and Process)

The NYSTP Grades 3–8 Common Core ELA and Mathematics Tests were administered in April 2014. The test items were selected from the pools of available ELA and mathematics items. These items were field-tested either in embedded field-testing or stand-alone field-testing in 2013.

The test construction process involved several iterative steps. Three criteria governed the item selection process. The first of these was to meet the ELA and mathematics content specifications provided by NYSED; the second was for content experts to select items with the best psychometric characteristics from the ELA and mathematics item pools, respectively; the third criteria required that the combined psychometric characteristics of all selected items combined to the intended psychometric goals for each entire form. Pearson content specialists were provided with the test designs, blueprints, and psychometric guidelines for item selection. The psychometric guidelines were based on the classical and item response theory (IRT) statistics associated with the test items. Using the pool of field-tested items, content specialists made preliminary selections for each grade and subject. The selections were then reviewed by the content leads for each subject to make sure items conformed to the different criteria. If the content criteria were not met, new items were selected. After the content leads' review, the item selections were reviewed by Pearson psychometricians. If items with undesirable statistics were

selected, the psychometricians proposed items with more desirable statistics and those items were reviewed by the content specialists and their leads. Once both Pearson content and psychometric teams were satisfied that the content and statistics of the selected items, and proposed whole forms met the requirements, the items were given to NYSED staff (including content and assessment experts) to review. Pearson content specialists and psychometricians traveled to Albany, New York, in September 2013 to finalize item selection and test creation with NYSED staff (including content and assessment experts). NYSED discussed the content and data of the proposed selections, explored alternate selections for consideration, approved the item selections, and assigned item positions to those items in the operational test books.

Test Form Production and Reviews

Once the selection of items for the operational and embedded field-test positions was completed, Pearson created test forms. The test forms were reviewed by Pearson content specialists and copyeditors and were posted for NYSED to review. NYSED and Pearson staff reviewed the forms looking for any spelling, capitalization, punctuation, grammar, and formatting errors. They also confirmed that each multiple-choice test item had a single correct answer.

Final Eyes Committee

After NYSED and Pearson staff reviewed copies of the test forms, the test forms were reviewed by the Final Eyes Committees. For each subject the committee consisted of approximately 12 Grade 3–8 New York State educators from around the state. During that review, educators were charged with taking the test to make sure that each multiple-choice item had a single correct answer and to look for spelling, capitalization, punctuation, grammar, and formatting errors. Upon completion of the Final Eyes review and after NYSED approved edits made as a result of the review, the tests were then considered final. The test files were then produced for the April 2014 administration.

Proficiency and Performance Standards

In summer 2013, after the operational administration of the 2013 tests, a standard-setting meeting occurred in Albany where 95 New York State educators went through a rigorous process guided by the best practices indicated by this intensely studied process to recommend performance standards for the new tests measuring the Common Core Learning Standards. These recommendations were presented to the Commissioner and the Board of Regents, who in turn adopted the recommended standards set forth by the committees. For additional details, see Section 8 and Appendix P in the 2013 technical report (NYSED, 2014).

For each grade level, there are four proficiency levels. Three cut points demarcate the performance standards needed to demonstrate each ascending level of proficiency. Detailed information related to performance standards can be found in Section 6, subsection, “Raw Score-to-Scale Score and SEM Conversion Tables.”

Section 3: Validity

Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of tests. Test validation is an ongoing process of gathering evidence from many sources to evaluate the soundness of the desired score interpretation or use. This evidence is acquired from studies of the content of the test as well as from studies involving scores produced by the test. Additionally, reliability has to be considered before considerations of validity are made. A test cannot be valid if the test scores are not also reliable.

The American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) *Standards for Educational and Psychological Testing* (1999) addressed the concept of validity in testing. Validity is the most important consideration in test evaluation. The concept refers to the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores. Test validation is the process for accumulating evidence to support any particular inference. Validity, however, is a unitary concept. Although evidence may be accumulated in many ways, validity refers to the degree to which evidence supports the inferences made from test scores.

Content Validity

Generally, achievement tests are used for student-level outcomes, either for making predictions about students or for describing students' performances (Mehrens and Lehmann, 1991). In addition, tests are now also used for the purpose of accountability and adequate yearly progress (AYP). NYSED uses various assessment data in reporting AYP. Specific to student-level outcomes, NYSTP documents student performance in the area of mathematics as defined by the New York State Common Core Mathematics Learning Standards and in the area of ELA as defined by the New York State Common Core ELA Learning Standards.

To allow test score interpretations appropriate for this purpose, the content of the test must be carefully matched to the specified standards. The 1999 AERA/APA/NCME standards state that content-related evidence of validity is a central concern during test development. Expert professional judgment should play an integral part in developing the definition of what is to be measured, such as describing the universe of the content, generating or selecting the content sample, and specifying the item format and scoring system.

Expert analysis of test content indicates the degree to which the content of a test covers the domain of content that the test is intended to measure. In the case of NYSTP, the content is defined by detailed blueprints that describe New York State content standards and that define the skills that must be measured to assess these content standards (see Tables B1 and B2 in Appendix B). The NYSTP test development process requires specific attention to content representation and the balance within each test form. New York State educators were involved in test construction in various development stages. For example, during the item review process, they reviewed field test items for the alignment of the items with the Common Core Learning Standards. Educators also participated in a process of establishing scoring rubrics (during Rangefinding sessions) for constructed-response items. Section 2, "Test Design and Development," contains more information specific to the item review process.

An external evaluation was conducted to review the development process and to investigate the degree to which NYSTP measures the CCLS. Human Resources Research Organization (HumRRO) found that NYSTP development process meets or exceeds the industry standards indicated in the *Standards for Educational and Psychological Testing* (1999). HumRRO also evaluated the degree to which the 2014 Grades 3–8 Common Core ELA and Mathematics Tests measured the CCLS. HumRRO concluded that the 2014 Grades 3–8 Common Core ELA and Mathematics Tests do assess the content described by the CCLS. Additionally, HumRRO found that NYSTP measured the CCLS at the intended Depth of Knowledge (DOK).

Construct (Internal Structure) Validity

Construct validity—what scores mean and what kind of inferences they support—is often considered the most important type of test validity. Construct validity of the NYSTP Grades 3–8 ELA and Mathematics Tests are supported by several types of evidence that can be obtained from the ELA and mathematics test data, respectively.

Internal Consistency

Empirical studies of the internal structure of the test provide one type of evidence of construct validity. For example, high internal consistency constitutes evidence of validity. This is because high coefficients imply that the test items are measuring the same domain of skill and are reliable and consistent. Reliability coefficients of the tests for total populations and subgroups of students are presented in Section 7 subsection, “Test Reliability.” For the total population, the ELA reliability coefficients (Cronbach’s alpha) ranged from .88 to .92, and for all subgroups, the reliability coefficients were greater than or equal to .80. For the total population, the mathematics reliability coefficients (Cronbach’s alpha) ranged from .93 to .95, and for all subgroups, the reliability coefficients were greater than or equal to .81. Overall, high internal consistency of the NYSTP Grades 3–8 Common Core ELA and Mathematics Tests provided sound evidence of construct validity.

Unidimensionality

Other validity evidence comes from analyses of the degree to which the test items conform to the requirements of the statistical models. These statistical models are used to scale and equate the tests, as well as to generate student scores. The models, among other things, require that the items fit the model well (item fit) and the items in a test measure a single domain of skill (unidimensionality).

The first step is to assess the degree to which the items fit the IRT model. The item-model fit for the mathematics and ELA tests was assessed using Q_I statistics (Yen, 1981), and the results are described in detail in Section 6, “IRT Calibration and Equating.” That the majority of the items demonstrated sound fit across grades and subjects, and only a few items were deemed to have deviate fit, provides solid evidence for the appropriateness of the IRT models used to calibrate and scale the test data.

Additional evidence for the efficacy of the model involves demonstrating that the items on New York State Tests are related to each other, within the respective subject areas. This relationship

of the items within the ELA or mathematics tests is, simply stated, the common proficiency acquired by students studying the content area. This “common proficiency,” or, more formally, underlying construct, could be labeled as ELA proficiency (using the ELA scores) or mathematics proficiency (using the mathematics scores), depending on the degree to which the ELA and mathematics items are related.

Factor analysis of the test data is one way of modeling the common construct. This analysis may show that there is a single or main factor that can account for much of the variability among responses to test items. A large first component in factor analysis would provide evidence of the latent proficiency students have in common with respect to the particular items asked. A large main factor found from a factor analysis of an achievement test would suggest a primary construct that may be related to what the items were designed to have in common (i.e., mathematics proficiency or ELA proficiency).

To demonstrate the common factor underlying student responses to ELA and mathematics test items, principal component factor analyses were conducted on a correlation matrix of individual items for the ELA and mathematics tests, respectively. Factoring a correlation (i.e., tetrachoric correlation) matrix rather than actual item response data is preferable when dichotomous variables are in the analyzed data set. Because the New York State ELA and mathematics tests contain both MC and CR items, the matrices of *polychoric* correlations were used as input for the factor analyses, as polychoric correlations are appropriate with MC and CR data. The study was conducted on the New York State public, charter, and non-public school students for whom data were available during the equating process. A large first principal component was evident in each analysis, demonstrating essential unidimensionality of the trait (i.e., proficiency) measured by each test. In other words, statistical evidence indicates the ELA items are measuring one underlying construct, ELA proficiency, and the mathematics items are measuring one underlying construct, mathematics proficiency.

The factor analyses conducted with the ELA and mathematics data will show almost as many underlying constructs, or factors, as there are items on the test. Therefore, it is necessary to further investigate the factor analysis results to determine the number of “meaningful” factors. Specifically, more than one factor with an eigenvalue greater than 1.0 present in each data set would suggest the presence of small additional factors. Additionally, the magnitude of the ratio of the variance accounted for by the first factor compared to the remaining factors also provides evidence as to the number of meaningful factors. In addition, the total amount of variance accounted for by the main factor was evaluated. According to M. Reckase (1979),

. . . the 1PL and the 3PL models estimate different abilities when a test measures independent factors, but . . . both estimate the first principal component when it is large relative to the other factors. In this latter case, good ability estimates can be obtained from the models, even when the first factor accounts for less than 10 percent of the test variance, although item calibration results will be unstable.

Factor analyses related to the Grades 3–8 Common Core ELA and Mathematics Tests indicated that the ratio of the variance accounted for by the first factor to the remaining factors was sufficiently large to support the claim that the ELA and mathematics tests were essentially

unidimensional; the ELA-related ratios and the mathematics-related ratios showed that the first eigenvalues were at least five times as large as the second eigenvalues for all the grades.

It was found that all the New York State Grades 3–8 Common Core ELA and Mathematics Tests exhibited first principal component accounting for more than 18% and 20% of the test variance, respectively. The results of factor analyses including eigenvalues greater than 1.0 and proportions of variance explained by the extracted factors are presented for ELA (see Table 4) and mathematics (see Table 5).

Table 4. ELA Tests Factor Analysis

Grade	Extracted Factor			
	#	Initial Eigenvalue	Variance Accounted for	
			%	Cumulative %
3	1	7.12	18.73	18.73
	2	1.43	3.77	22.50
	3	1.21	3.17	25.67
4	1	7.99	19.48	19.48
	2	1.64	4.01	23.49
	3	1.12	2.73	26.22
	4	1.04	2.53	28.75
	5	1.02	2.49	31.24
5	1	9.64	18.53	18.53
	2	1.65	3.16	21.69
	3	1.27	2.45	24.14
	4	1.08	2.08	26.22
	5	1.04	1.99	28.21
	6	1.01	1.93	30.14
6	1	10.52	20.24	20.24
	2	1.78	3.42	23.66
	3	1.35	2.60	26.26
	4	1.10	2.12	28.38
	5	1.04	2.00	30.38
7	1	9.53	18.69	18.69
	2	1.71	3.35	22.04
	3	1.28	2.51	24.55
	4	1.10	2.16	26.71
	5	1.06	2.07	28.78
	6	1.01	1.98	30.76
8	1	9.91	19.06	19.06
	2	1.81	3.49	22.55
	3	1.23	2.37	24.92
	4	1.10	2.12	27.04
	5	1.05	2.02	29.06
	6	1.01	1.94	31.00

This evidence supports the claim that there is one single construct underlying the items/tasks in each ELA test and that scores from each test would represent performance primarily determined by that construct. Construct-irrelevant variance does not appear to create significant nuisance factors.

Table 5. Mathematics Tests Factor Analysis

Grade	Extracted Factor			
	#	Initial Eigenvalue	Variance Accounted for	
			%	Cumulative %
3	1	11.77	24.01	24.01
	2	1.89	3.85	27.86
	3	1.17	2.39	30.25
	4	1.06	2.17	32.42
	5	1.00	2.05	34.47
4	1	14.71	28.28	28.28
	2	1.50	2.88	31.16
	3	1.21	2.32	33.48
	4	1.08	2.08	35.56
	5	1.06	2.04	37.60
5	1	13.04	25.08	25.08
	2	1.93	3.70	28.78
	3	1.29	2.48	31.26
	4	1.05	2.01	33.27
	5	1.01	1.93	35.20
6	1	14.14	24.38	24.38
	2	2.03	3.49	27.87
	3	1.27	2.19	30.06
	4	1.06	1.83	31.89
	5	1.04	1.80	33.69
7	1	14.15	24.40	24.40
	2	1.96	3.38	27.78
	3	1.24	2.14	29.92
	4	1.01	1.74	31.66
8	1	11.78	20.31	20.31
	2	1.58	2.72	23.03
	3	1.18	2.03	25.06
	4	1.07	1.85	26.91
	5	1.03	1.77	28.68

This evidence supports the claim that there is a common construct underlying the items/tasks in each mathematics test and that scores from each test would represent performance primarily

determined by that construct. Construct-irrelevant variance does not appear to create significant nuisance factors.

As additional evidence for construct validity, the same factor analysis procedure was employed to assess the dimensionality of the mathematics construct for selected subgroups of students in each grade: English language learners (ELL), students with disabilities (SWD), and students using test accommodations (SUA). The results were comparable to the results obtained from the total population data. Evaluation of eigenvalue magnitude and proportions of variance explained by the main and secondary factors provide evidence of essential unidimensionality of the construct measured by the tests for the analyzed subgroups. Factor analysis results for ELL, SWD, SUA, ELL/SUA, and SWD/SUA classifications are provided in Appendix L. The ELL/SUA subgroup is defined as examinees who are English language learners and who use at least one ELL-related accommodation. The SWD/SUA subgroup includes examinees that are classified with disabilities that use at least one disability-related accommodation.

Detection of Bias

Minimizing item bias has the goal of minimizing construct-irrelevant variance and helps establish a strong validity argument for the tests. Specifically, bias occurs if items function differentially for key pairs of groups, which may in turn cause the test to be differentially valid for certain groups of test takers. The statistical means for flagging items that may exhibit bias is referred to as differential item functioning (DIF) and these statistical procedures were designed to be conservative—in other words—to flag more items for DIF rather than fewer. So it is rare in practice to observe a high stakes test in which not a single item is flagged for DIF. And since those procedures tend to over-flag items, it is only through review by experts of those flagged items that the items flagged for DIF may be judged to have or be free of bias. If the test involves irrelevant skills or knowledge, the possibility of bias is increased. Thus, preserving content validity is essential.

The developers of the NYSTP tests gave careful attention to items of possible ethnic, gender, socioeconomic status (SES), and—only for the mathematics tests—translation bias. All materials were written and reviewed to conform to Pearson’s editorial policies and guidelines for equitable assessment, as well as NYSED’s guidelines for item development. At the same time, all materials were written to NYSED’s specifications and carefully checked by groups of trained New York State educators during the item review process. These steps are essential in keeping bias to a minimum. However, current evidence suggests that expertise in this area is no substitute for data; reviewers are sometimes wrong about which items work to the disadvantage of a group, apparently because some of their ideas about how students will react to items may be faulty (Sandoval and Mille, 1979; Jensen, 1980). Thus, empirical studies were conducted.

Statistical methods were used to identify items exhibiting possible DIF. Although items flagged for DIF in the field test stage were closely examined for content bias and avoided during the operational test construction, DIF analyses were conducted again on operational test data. Different methods were employed to evaluate the amount of DIF in all test items: constructed-response items were evaluated with standardized mean differences, and multiple-choice items were analyzed using Mantel-Haenszel methods (see Section 5, “Operational Test Data Collection and Classical Analysis”).

In each grade for both the ELA and mathematics tests, few items were flagged for DIF. Moreover, the magnitude of DIF for the flagged items was typically small (for more detail, see Appendix N). In addition, very few items were flagged by multiple methods. Items that were flagged for statistically significant DIF were carefully reviewed by multiple reviewers during the operational test item selection. All such items were deemed by the reviewers to be free of bias (i.e., judged not to adversely impact any demographic subgroup studied) and remained in the tests.

Section 4: Test Administration and Scoring

Listed in this section are brief summaries of New York State test administration and scoring procedures. For further information, refer to the aforementioned School Administrator’s Manual and the *New York State Scoring Leader Handbook (2014)* located here:

<http://www.p12.nysed.gov/assessment/sam/ei/scoringleaderhandbook14.pdf>.

Test Administration

NYSTP Grades 3–8 Common Core ELA and Mathematics Tests were administered to students during April 2014. The testing window was April 1–3 for the Grades 3–8 Common Core ELA Tests and April 30–May 2 for the Grades 3–8 Common Core Mathematics Tests. The makeup test administration window was April 4, 7, and 8 for the Grades 3–8 Common Core ELA Tests and May 5–7 for the Grades 3–8 Common Core Mathematics Tests. The makeup test administration windows allowed students who were ill or otherwise unable to test during the assigned window to take the tests.

Scoring Procedures of Operational Tests

The scoring of the Common Core operational tests was performed at designated sites by qualified teachers and administrators. The number of personnel at a given site varied, as districts have the option of regional, district-wide, or school-wide scoring (please refer to the next subsection, “Scoring Models,” for more detail). Administrators were responsible for the oversight of scoring operations, including the preparation of the test site, the security of test books, and the supervision of the scoring process. At each site, designated trainers taught scoring committee members the basic criteria for scoring each item and monitored the scoring sessions in the room. The trainers were assisted by facilitators or leaders who also helped in monitoring the sessions and enforced scoring accuracy.

The titles for administrators, trainers, and facilitators vary by the scoring model that is selected. At the regional level, oversight was conducted by a site coordinator. A scoring leader trained the scoring committee members and monitored the sessions, and a table facilitator assisted in monitoring the sessions. For each subject, the oversight was structured in the same way for district- and school-wide models. At the district-wide level, a school district administrator oversaw scoring. A district subject leader trained the scoring committee members and monitored the sessions, and a school subject leader assisted in monitoring the sessions. For school-wide scoring, oversight was provided by the principal; otherwise, titles for the school-wide model were the same as those for the district-wide model. The general title “scoring-committee members” included scorers at every site.

Scoring Models

For the 2013–2014 school year, schools and school districts were able to score Grades 3–8 Common Core ELA and/or Mathematics Tests regionally, multi-district, district-wide, or school-wide based on local need. Schools were required to enter one of the following scoring model codes on student answer sheets:

1. Regional scoring—The scorers for the school’s test papers included either staff from three or more school districts or staff from all non-public schools in an affiliation group (non-public or charter schools may participate in regional scoring with public school districts and may be counted as one district).
2. Schools from two districts—The scorers for the school’s test papers included staff from two school districts, non-public schools, charter school districts, or a combination thereof.
3. Three or more schools within a district—The scorers for the school’s test papers included staff from all schools administering this test in a district, provided at least three schools are represented.
4. Two schools within a district—The scorers for the school’s test papers included staff from all schools administering this test in a district, provided that two schools are represented.
5. One school only (local scoring)—The first readers for the school’s test papers included staff from the only school in the district administering this test, staff from one charter school, or staff from one non-public school.

Schools and districts were instructed to carefully analyze their individual needs and capacities to determine their appropriate scoring model. Boards of Cooperative Educational Services (BOCES) and the Staff and Curriculum Development Network (SCDN) provided districts with technical support and advice in making this decision.

Scoring of Constructed-Response Items

The key resource for both the training of scoring committee members and the scoring of CR items were the scoring guides. These documents were created by Pearson from sets of actual field test student responses that were consensus scored by NYSED and New York State teachers during Rangefinding sessions. Trainers used these materials to train scoring-committee members on the criteria for scoring CR items. Additionally, scoring leader handbooks were also distributed to outline the responsibilities of the scoring roles. Pearson and NYSED staff also conducted turnkey training sessions across the state to better equip the teachers and administrators with enhanced knowledge of scoring principles and criteria.

Upon completion of the training of scoring committee members, scoring was conducted with pen-and-pencil scoring as opposed to electronic scoring, and each scoring-committee member evaluated actual student papers instead of electronically scanned papers. All scoring-committee members were trained by previously trained and approved trainers along with guidance from scoring guides. Each constructed-response test book was scored by three separate scoring committee members, who scored three distinct sections of the test book. After test books were completed, the table facilitator or subject (ELA or mathematics) leader conducted a “read

behind” of approximately 12 sets of test books per hour to verify the accuracy of scoring. If an item arose that was not covered in the training materials, facilitators or trainers were to call the New York State ELA Helpline or the New York State Mathematics Helpline for assistance with the ELA or mathematics scoring, respectively (see the subsection “Quality Control Process”).

Scorer Qualifications and Training

The scoring of the Common Core operational tests was conducted by qualified administrators and teachers. Trainers used the scoring guides to train scoring-committee members on the criteria for scoring constructed-response items. Part of the training process was the administration of a consistency assurance set (CAS) that provided the state’s scoring sites with information regarding strengths and weaknesses of their scorers. This tool allowed trainers to retrain their scorers, if necessary. The CAS also acknowledged those scorers who had grasped all aspects of the content area being scored and was well prepared to score student responses.

Regardless of the scoring model used, a minimum of three scorers is necessary to score each student’s test. However, to comply with a State requirement, none of the scorers assigned to score a student’s test responses may be that student’s teacher. This policy is detailed in the *School Administrator’s Manual* section “Assigning Scorer Numbers and Questions to Scoring Committee Members” on page 22, found online at:

<http://www.p12.nysed.gov/assessment/sam/ei/scoringleaderhandbook14.pdf>.

Quality Control Process

Test books were randomly distributed throughout each scoring room so that books from each region, district, school, or class were evenly dispersed. Teams were divided into groups of three to ensure that a variety of scorers graded each book. If a scorer and a facilitator could not reach a decision on a paper after reviewing the scoring guides and audio files, they called the New York State ELA or Mathematics Helpline. The call center was established to help teachers and administrators during scoring. The help-line staff consisted of trained Pearson personnel who answered items by phone or fax. When a member of the staff was unable to resolve an issue, it was referred to NYSED for a scoring decision. A quality check was also performed on each completed box of scored tests to certify that all items were scored and that the scoring-committee members darkened each score on the answer document appropriately. The log of calls received by the scoring helpline was delivered to NYSED twice daily during the scoring window. To affirm that all schools across the state adhered to scoring guidelines and policies, approximately 5% of the schools’ results are audited each year by an outside vendor.

Section 5: Operational Test Data Collection and Classical Analysis

Data Collection

Test data were collected in two phases. During Phase 1, a sample of approximately 96% of the student test records were received from the data warehouse and delivered to Pearson, beginning at the end of May 2014. Phase 2 involved submitting “straggler files” to Pearson in June 2014.

The straggler files contained less than 2% of the total population cases and were excluded from the classical, IRT, and reliability analyses (as described in Sections 5, 6, and 7, respectively) due to late submission. The analyses described in Section 8, “Summary of Operational Test Results” were based on the data collected from both Phase 1 and Phase 2. Data collected from both public schools and non-public schools were included in all data analyses.

Data Processing

Depending on the nature of the analysis, more student records were included in some analyses than in others. For example, all students with a valid test score were included in the analyses described in Section 8, “Summary of Operational Test Results.” For the analyses described in other sections, however, more stringent data cleaning procedures were applied (see details below).

Data processing here refers to the cleaning and screening procedures used to identify errors (such as out-of-range data), and the decisions made to exclude student cases or to suppress particular items in certain analyses. Pearson’s psychometric team performed data cleaning to the delivered data and excluded some student cases in order to obtain a sample of the utmost integrity. It should be noted that a student case being excluded from certain data analyses did not mean that the student record was invalidated. According to the NYSED’s specific instructions, additional procedures were taken to correct or recover these student records so that their test results were scored properly. As mentioned above, their records were included in Section 8 analyses.

The major groups of cases excluded from the data set (used for analyses in Sections 5, 6, and 7) were students with missing school type and those with at least one entirely missing test book. Other deleted cases included students with incorrect or incomplete grade information; duplicate record cases; and no-response record cases. The mathematical data cleaning procedure also excluded records with mismatched form language indicators for translated versions across the three test books for a given student. The data cleaning procedures and accompanying case counts are represented for ELA and mathematics in Table 6A–6F and Table 7A–7F, respectively.

One additional data cleaning rule was implemented for the Grade 3 mathematics test. Due to a printing error in Book 2 of one of the four operational forms, items in positions (i.e., actual tested sequences) 34–37 and 44–48 may have been missing from this book. The items were all multiple-choice, 7 of which were operational (including one internal anchor item) and two of which were field test positions (including one external anchor item). As soon as the issue came to light, the New York State Education Department made every effort to notify affected schools and assist them in obtaining copies of the affected pages; either by sending new test books, if time permitted, or by faxing or electronically submitting secure copies of the relevant pages, if

time did not permit. Despite the fact that these efforts are believed to have resulted in remedies for all affected students, additional data cleaning rules were put into place.

Two scenarios were considered, the first was one in which students simply skipped the missing nine items and the second was meant to capture those students who—not noticing the missing item numbers—responded in order, leaving the last nine items blank and resulting in a shifted answer string for items 35 through the end of Book 2. First, students must have taken the relevant form for test book 2 without translation (i.e., they tested in English). To address the omission scenario, students were flagged if they had at least one response in the combined sequence range 25–33 and 38–43 and they omitted all items in sequence ranges 34–37 and 44–48. To address the issue of the shifted response string, flagged students must have had at least one response in the sequence range 25–39 and they must have omitted all items in sequence range 40–48 (i.e., the last nine items). In addition to having removed these flagged students from the equating sample, their test results were deemed mis-administrations, so as to inadvertently disadvantage neither the student nor his or her school. The result—shown in Table 7A in the row labeled “Flagged for Printing Issue”—was that only 19 records were removed from the equating sample (*n-count* = 200,278).

Table 6A. ELA Grade 3 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	200,195
Wrong Subject	0	200,195
No Grade	113	200,082
Wrong Grade	73	200,009
Language Mismatched Form	5	200,004
School Type	462	199,542
Missing Entire Book	1,042	198,500
Invalid Score	0	198,500
Out-of-Range CR Scores	0	198,500
Duplicated Record	12	198,488

Table 6B. ELA Grade 4 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	205,093
Wrong Subject	0	205,093
No Grade	79	205,014
Wrong Grade	104	204,910
Language Mismatched Form	3	204,907
School Type	571	204,336
Missing Entire Book	526	203,810
Invalid Score	0	203,810
Out-of-Range CR Scores	0	203,810
Duplicated Record	10	203,800

Table 6C. ELA Grade 5 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	200,767
Wrong Subject	0	200,767
No Grade	168	200,599
Wrong Grade	21	200,578
Language Mismatched Form	1	200,577
School Type	577	200,000
Missing Entire Book	571	199,429
Invalid Score	0	199,429
Out-of-Range CR Scores	0	199,429
Duplicated Record	2	199,427

Table 6D. ELA Grade 6 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	202,044
Wrong Subject	0	202,044
No Grade	96	201,948
Wrong Grade	118	201,830
Language Mismatched Form	3	201,827
School Type	666	201,161
Missing Entire Book	717	200,444
Invalid Score	0	200,444
Out-of-Range CR Scores	0	200,444
Duplicated Record	4	200,440

Table 6E. ELA Grade 7 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	199,803
Wrong Subject	0	199,803
No Grade	92	199,711
Wrong Grade	162	199,549
Language Mismatched Form	2	199,547
School Type	935	198,612
Missing Entire Book	1,003	197,609
Invalid Score	0	197,609
Out-of-Range CR Scores	0	197,609
Duplicated Record	6	197,603

Table 6F. ELA Grade 8 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	206,014
Wrong Subject	0	206,014
No Grade	102	205,912
Wrong Grade	92	205,820
Language Mismatched Form	5	205,815
School Type	1,238	204,577
Missing Entire Book	1,217	203,360
Invalid Score	0	203,360
Out-of-Range CR Scores	0	203,360
Duplicated Record	2	203,358

Table 7A. Mathematics Grade 3 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	201,183
Wrong Subject	0	201,183
No Grade	105	201,078
Wrong Grade	77	201,001
Language Mismatched Form	17	200,984
School Type	442	200,542
Missing Entire Book	233	200,309
Invalid Score	0	200,309
Out-of-Range CR Scores	0	200,309
Duplicated Record	12	200,297
Flagged for Printing Issue	19	200,278

Table 7B. Mathematics Grade 4 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	205,802
Wrong Subject	0	205,802
No Grade	82	205,720
Wrong Grade	116	205,604
Language Mismatched Form	64	205,540
School Type	562	204,978
Missing Entire Book	200	204,778
Invalid Score	0	204,778
Out-of-Range CR Scores	0	204,778
Duplicated Record	10	204,768

Table 7C. Mathematics Grade 5 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	200,407
Wrong Subject	0	200,407
No Grade	173	200,234
Wrong Grade	22	200,212
Language Mismatched Form	51	200,161
School Type	564	199,597
Missing Entire Book	1,023	198,574
Invalid Score	0	198,574
Out-of-Range CR Scores	0	198,574
Duplicated Record	4	198,570

Table 7D. Mathematics Grade 6 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	200,437
Wrong Subject	0	200,437
No Grade	92	200,345
Wrong Grade	133	200,212
Language Mismatched Form	49	200,163
School Type	648	199,515
Missing Entire Book	381	199,134
Invalid Score	0	199,134
Out-of-Range CR Scores	0	199,134
Duplicated Record	4	199,130

Table 7E. Mathematics Grade 7 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	195,229
Wrong Subject	0	195,229
No Grade	82	195,147
Wrong Grade	156	194,991
Language Mismatched Form	75	194,916
School Type	922	193,994
Missing Entire Book	645	193,349
Invalid Score	0	193,349
Out-of-Range CR Scores	0	193,349
Duplicated Record	6	193,343

Table 7F. Mathematics Grade 8 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	161,522
Wrong Subject	0	161,522
No Grade	97	161,425
Wrong Grade	85	161,340
Language Mismatched Form	63	161,277
School Type	1,189	160,088
Missing Entire Book	776	159,312
Invalid Score	0	159,312
Out-of-Range CR Scores	0	159,312
Duplicated Record	8	159,304

Classical Analysis and Calibration Sample Characteristics

The demographic characteristics of students in the cleaned calibration and equating data sets are presented in the following tables, with the ELA tables provided first (Table 8A–8F), followed by the mathematics tables (Table 9A–9F). The clean data sets included over 96% of New York State students and were used for classical analyses presented in the calibrations in this section. The Needs/Resource Capacity Category (NRC) is assigned at the district level and is an indicator of district and school socioeconomic status. The ethnicity and gender designations are based on student level information.

Table 8A. ELA Grade 3 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	97,653	49.20
	Male	100,835	50.80
Ethnicity	Asian	17,134	8.63
	Black	35,369	17.82
	Hispanic	50,691	25.54
	American Indian	1,161	0.58
	Multiracial	3,590	1.81
	Pacific Islander	511	0.26
	White	90,032	45.36
NRC	New York	70,913	35.76
	Big 4 Cities	7,737	3.90
	Urban/Suburban	15,030	7.58
	Rural	10,327	5.21
	Average Needs	51,370	25.91
	Low Needs	24,137	12.17
	Charter	8,310	4.19
	Non-Public	10,453	5.27
SWD	No	169,790	85.54
	Yes	28,698	14.46
SUA	No	174,582	87.96
	Yes	23,906	12.04
ELL	No	182,905	92.15
	Yes	15,583	7.85

Note. The sample n-count was 198,488.

Table 8B. ELA Grade 4 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	100,639	49.38
	Male	103,161	50.62
Ethnicity	Asian	17,948	8.81
	Black	35,717	17.53
	Hispanic	49,427	24.25
	American Indian	1,125	0.55
	Multiracial	2,960	1.45
	Pacific Islander	456	0.22
	White	96,167	47.19
NRC	New York	69,248	34.01
	Big 4 Cities	7,741	3.80
	Urban/Suburban	14,505	7.12
	Rural	10,381	5.10
	Average Needs	52,501	25.79
	Low Needs	25,985	12.76
	Charter	7,065	3.47
	Non-Public	16,166	7.94
SWD	No	173,385	85.08
	Yes	30,415	14.92
SUA	No	177,386	87.04
	Yes	26,414	12.96
ELL	No	187,539	92.02
	Yes	16,261	7.98

Note. The sample n-count was 203,800.

Table 8C. ELA Grade 5 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	98,168	49.23
	Male	101,259	50.77
Ethnicity	Asian	17,219	8.63
	Black	35,520	17.81
	Hispanic	47,929	24.03
	American Indian	1,138	0.57
	Multiracial	2,540	1.27
	Pacific Islander	457	0.23
	White	94,624	47.45
NRC	New York	67,208	33.74
	Big 4 Cities	7,642	3.84
	Urban/Suburban	14,478	7.27
	Rural	10,458	5.25
	Average Needs	54,036	27.13
	Low Needs	26,634	13.37
	Charter	8,123	4.08
	Non-Public	10,617	5.33
SWD	No	168,026	84.25
	Yes	31,401	15.75
SUA	No	173,912	87.21
	Yes	25,515	12.79
ELL	No	186,231	93.38
	Yes	13,196	6.62

Note. The sample n-count was 199,427.

Table 8D. ELA Grade 6 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	99,117	49.45
	Male	101,323	50.55
Ethnicity	Asian	16,959	8.46
	Black	37,103	18.51
	Hispanic	48,052	23.97
	American Indian	1,112	0.55
	Multiracial	2,337	1.17
	Pacific Islander	407	0.20
	White	94,470	47.13
NRC	New York	65,580	32.75
	Big 4 Cities	7,448	3.72
	Urban/Suburban	14,171	7.08
	Rural	10,497	5.24
	Average Needs	52,821	26.38
	Low Needs	26,484	13.23
	Charter	8,170	4.08
	Non-Public	15,076	7.53
SWD	No	169,763	84.70
	Yes	30,677	15.30
SUA	No	174,546	87.08
	Yes	25,894	12.92
ELL	No	187,158	93.37
	Yes	13,282	6.63

Note. The sample n-count was 200,440.

Table 8E. ELA Grade 7 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	97,174	49.18
	Male	100,429	50.82
Ethnicity	Asian	16,496	8.35
	Black	36,896	18.67
	Hispanic	46,537	23.55
	American Indian	1,096	0.55
	Multiracial	2,030	1.03
	Pacific Islander	388	0.20
	White	94,160	47.65
NRC	New York	66,593	33.74
	Big 4 Cities	7,480	3.79
	Urban/Suburban	14,166	7.18
	Rural	10,988	5.57
	Average Needs	53,091	26.90
	Low Needs	28,019	14.20
	Charter	6,441	3.26
	Non-Public	10,603	5.37
SWD	No	167,240	84.63
	Yes	30,363	15.37
SUA	No	176,242	89.19
	Yes	21,361	10.81
ELL	No	186,736	94.50
	Yes	10,867	5.50

Note. The sample n-count was 197,603.

Table 8F. ELA Grade 8 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	99,644	49.00
	Male	103,714	51.00
Ethnicity	Asian	17,883	8.79
	Black	38,059	18.72
	Hispanic	47,207	23.21
	American Indian	1,073	0.53
	Multiracial	1,924	0.95
	Pacific Islander	432	0.21
	White	96,780	47.59
NRC	New York	68,859	33.89
	Big 4 Cities	7,201	3.54
	Urban/Suburban	13,806	6.80
	Rural	10,758	5.30
	Average Needs	53,027	26.10
	Low Needs	28,632	14.09
	Charter	5,061	2.49
	Non-Public	15,810	7.78
SWD	No	174,039	85.58
	Yes	29,319	14.42
SUA	No	181,522	89.26
	Yes	21,836	10.74
ELL	No	192,656	94.74
	Yes	10,702	5.26

Note. The sample n-count was 203,358.

Table 9A. Mathematics Grade 3 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	98,086	48.97
	Male	102,192	51.03
Ethnicity	Asian	17,655	8.82
	Black	35,570	17.76
	Hispanic	51,852	25.89
	American Indian	1,162	0.58
	Multiracial	3,594	1.79
	Pacific Islander	520	0.26
	White	89,925	44.90
NRC	New York	72,379	36.18
	Big 4 Cities	7,920	3.96
	Urban/Suburban	15,202	7.60
	Rural	10,283	5.14
	Average Needs	51,149	25.57
	Low Needs	24,253	12.12
	Charter	8,320	4.16
	Non-Public	10,561	5.28
SWD	No	171,011	85.39
	Yes	29,267	14.61
SUA	No	167,028	83.40
	Yes	33,250	16.60
ELL	No	182,546	91.15
	Yes	17,732	8.85

Note. The sample n-count was 200,297.

Table 9B. Mathematics Grade 4 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	100,866	49.26
	Male	103,902	50.74
Ethnicity	Asian	18,401	8.99
	Black	35,799	17.48
	Hispanic	50,242	24.54
	American Indian	1,126	0.55
	Multiracial	2,964	1.45
	Pacific Islander	466	0.23
	White	95,770	46.77
NRC	New York	70,436	34.43
	Big 4 Cities	7,895	3.86
	Urban/Suburban	14,638	7.16
	Rural	10,323	5.05
	Average Needs	52,112	25.48
	Low Needs	25,904	12.66
	Charter	7,057	3.45
	Non-Public	16,192	7.92
SWD	No	174,123	85.03
	Yes	30,645	14.97
SUA	No	169,033	82.55
	Yes	35,735	17.45
ELL	No	186,450	91.05
	Yes	18,318	8.95

Note. The sample n-count was 204,768.

Table 9C. Mathematics Grade 5 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	97,562	49.13
	Male	101,008	50.87
Ethnicity	Asian	17,607	8.87
	Black	35,284	17.77
	Hispanic	48,312	24.33
	American Indian	1,123	0.57
	Multiracial	2,536	1.28
	Pacific Islander	459	0.23
	White	93,249	46.96
NRC	New York	68,327	34.45
	Big 4 Cities	7,712	3.89
	Urban/Suburban	14,539	7.33
	Rural	10,314	5.20
	Average Needs	52,154	26.29
	Low Needs	26,318	13.27
	Charter	8,130	4.10
	Non-Public	10,852	5.47
SWD	No	167,337	84.27
	Yes	31,233	15.73
SUA	No	163,386	82.28
	Yes	35,184	17.72
ELL	No	183,463	92.39
	Yes	15,107	7.61

Note. The sample n-count was 198,570.

Table 9D. Mathematics Grade 6 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	98,050	49.24
	Male	101,080	50.76
Ethnicity	Asian	17,337	8.71
	Black	36,931	18.55
	Hispanic	48,743	24.48
	American Indian	1,076	0.54
	Multiracial	2,288	1.15
	Pacific Islander	433	0.22
	White	92,322	46.36
NRC	New York	66,787	33.57
	Big 4 Cities	7,543	3.79
	Urban/Suburban	14,154	7.11
	Rural	10,257	5.16
	Average Needs	51,112	25.69
	Low Needs	25,827	12.98
	Charter	8,150	4.10
	Non-Public	15,121	7.60
SWD	No	168,809	84.77
	Yes	30,321	15.23
SUA	No	165,709	83.22
	Yes	33,421	16.78
ELL	No	183,877	92.34
	Yes	15,253	7.66

Note. The sample n-count was 199,130.

Table 9E. Mathematics Grade 7 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	94,663	48.96
	Male	98,680	51.04
Ethnicity	Asian	16,670	8.62
	Black	36,433	18.84
	Hispanic	46,771	24.19
	American Indian	1,054	0.55
	Multiracial	1,932	1.00
	Pacific Islander	391	0.20
	White	90,092	46.60
NRC	New York	67,561	34.98
	Big 4 Cities	7,486	3.88
	Urban/Suburban	13,987	7.24
	Rural	10,529	5.45
	Average Needs	50,098	25.94
	Low Needs	26,445	13.69
	Charter	6,415	3.32
	Non-Public	10,616	5.50
SWD	No	163,772	84.71
	Yes	29,571	15.29
SUA	No	168,893	87.35
	Yes	24,450	12.65
ELL	No	180,521	93.37
	Yes	12,822	6.63

Note. The sample n-count was 193,343.

Table 9F. Mathematics Grade 8 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	76,750	48.18
	Male	82,554	51.82
Ethnicity	Asian	12,040	7.56
	Black	33,981	21.33
	Hispanic	42,623	26.76
	American Indian	892	0.56
	Multiracial	1,446	0.91
	Pacific Islander	321	0.20
	White	68,001	42.69
NRC	New York	59,190	37.18
	Big 4 Cities	6,812	4.28
	Urban/Suburban	12,110	7.61
	Rural	8,872	5.57
	Average Needs	36,583	22.98
	Low Needs	15,943	10.02
	Charter	4,861	3.05
	Non-Public	14,808	9.30
SWD	No	131,986	82.85
	Yes	27,318	17.15
SUA	No	133,779	83.98
	Yes	25,525	16.02
ELL	No	146,867	92.19
	Yes	12,437	7.81

Note. The sample n-count was 159,304.

Classical Data Analysis

Classical data analysis of the NYSTP Grades 3–8 ELA and Mathematics Tests consists of several important elements. One element is the analysis of item-level statistical information about student performance. It is important to verify that the items and test forms function as intended. If any serious error were to occur with an item (i.e., a printing error or two correct answers to one item), item analysis is the stage at which errors should be flagged and evaluated for rectification (suppression, credit, or other acceptable solution). Analyses of test-level data comprise the second element of classical data analysis. These include examination of the raw score (RS) statistics (mean and standard deviation or “SD”) and test reliability measures Cronbach’s alpha (Cronbach, 1951) and Feldt-Raju coefficient (Qualls, 1995). Assessment of test speededness is another important element of classical analysis. Additionally, classical DIF analysis is conducted at this stage. DIF analysis includes computation of standardized mean differences and Mantel-Haenszel statistics for New York State items to identify potential item bias. All classical data analysis results contribute information on the validity and reliability of the tests (also see Section 3, “Validity,” and Section 7, “Reliability and Standard Error of Measurement”).

Item Difficulty and Point Biserial Correlation Coefficients

Item difficulty is classically measured by the p-value statistic. It assesses the proportion of students who responded correctly to each MC item or the average proportion of the maximum score that students earned on each CR item. It is important to have a good range of p-values to increase test information and to avoid floor or ceiling effects. P-values represent the overall degree of difficulty, but do not account for demonstrated student performance on other test items. Usually, p-value information is coupled with point biserial (pbis) statistics, to verify that items are functioning as intended. In Appendix M, Tables M1–M12 illustrate classical test statistics for all items on each grade level test. Appendix F provides general psychometric guidelines for operational item selection.

Item difficulties (p-values) on the ELA tests ranged from .25 to .92. For Grade 3, the item p-values ranged from .32 to .81, with a mean of .53. For Grade 4, the item p-values ranged from .25 to .77, with a mean of .57. For Grade 5, the item p-values ranged from .26 to .90, with a mean of .59. For Grade 6, the item p-values ranged from .36 to .92, with a mean of .60. For Grade 7, the item p-values ranged from .28 to .81, with a mean of .56. For Grade 8, the item p-values ranged from .40 to .91, with a mean of .63. These p-value statistics are in Appendix M Tables M1–M6, along with pbis statistics of the keys.

Item difficulties (p-values) on the mathematics tests ranged from .19 to .96. For Grade 3, the item p-values ranged from .32 to .96, with a mean of .66. For Grade 4, the item p-values ranged from .40 to .92, with a mean of .63. For Grade 5, the item p-values ranged from .20 to .90, with a mean of .61. For Grade 6, the item p-values ranged from .19 to .91, with a mean of .60. For Grade 7, the item p-values ranged from .22 to .84, with a mean of .54. For Grade 8, the item p-values ranged from .20 to .89, with a mean of .53. These statistics are provided in Appendix M Tables M7–M12, along with other classical test summary statistics.

Point biserial statistics are used to examine item-test correlations or item discrimination for MC items. The pbis correlation for the key (i.e., the correct answer) is a measure of internal consistency, while pbis for specific response options aid in flagging possible alternate keys; each is a correlation which ranges between ± 1 . It is the correlation of students' responses to an item relative to their performance on the rest of the test and, unless otherwise noted, this discussion will be limited to the point biserial of the correct response with the remainder of the test.

Point biserial correlations are presented in Appendix M Tables M1–M12. The column labeled “Pbis Key” contains the point biserial correlation associated with the correct response. The guideline for building the NYSTP Grades 3–8 Common Core ELA and Mathematics Tests was that the pbis correlation for the key for MC items should be equal to or greater than .20, which would indicate that students who responded correctly to that item also tended to do well on the overall test. There were very few exceptions to this guideline due to content considerations, which required the inclusion of particular items. Decisions to use such items were made very carefully, and no item with a negative point-biserial correlation was allowed on the test.

Point biserials for correct answer options on the ELA tests ranged from .10 to .56, as shown in Appendix M in Tables M1–M6. For Grade 3, the item pbis values ranged from .23 to .50, with a mean of .37. For Grade 4, the item pbis values ranged from .13 to .51, with a mean of .35. For

Grade 5, the item pbis values ranged from .10 to .52, with a mean of .37. For Grade 6, the item pbis values ranged from .26 to .56, with a mean of .40. For Grade 7, the item pbis values ranged from .22 to .55, with a mean of .36. For Grade 8, the item pbis values ranged from .24 to .52, with a mean of .38.

Point biserial correlations for the correct answer options on the mathematics tests ranged from .15 to .66, as shown in Appendix M in Tables M7–M12. For Grade 3, the item pbis values ranged from .23 to .63, with a mean of .45. For Grade 4, the item pbis values ranged from .33 to .66, with a mean of .48. For Grade 5, the item pbis values ranged from .24 to .63, with a mean of .46. For Grade 6, the item pbis values ranged from .29 to .64, with a mean of .45. For Grade 7, the item pbis values ranged from .23 to .59, with a mean of .43. For Grade 8, the item pbis values ranged from .15 to .52, with a mean of .39.

Speededness

Speededness refers to interference in test scores due to insufficient testing time. It is NYSED policy that ample testing time should be given for students to complete the entire test. Furthermore, both the validity (i.e., accuracy) and reliability (i.e., precision) of test scores are adversely affected when tests are speeded. For these reasons, sufficient administration time limits were set for the NYSTP tests.

Speededness is routinely checked, based on test data, after each administration. One method of analyzing data to determine if speededness has occurred is to review the proportion of students not answering (i.e., omitting) items, especially those items that appeared towards the end of the test form. Tables M1–M12 in Appendix M show the omit rates for items on the Grades 3–8 Common Core ELA and Mathematics Tests, respectively. The industry standard general rule of thumb is that omit rates for multiple-choice items should be less than 5.0%. Omit rates across multiple-choice and constructed-response items on the Grades 3–8 Common Core ELA and Mathematics Tests typically ranged from 0% to 3%. As may be expected, omit rates tended to increase for items at the end of the test booklets, and only for ELA Grade 3 in Books 1 and 3 did items initially exceed that 5% threshold. It was for that reason that the last two operational items in Book 1 (both MC) and the last operational item in Book 3 (a 4-point CR item) were dropped from scoring and all analyses presented herein. In general, omit rates rarely exceeded 3%, even for the last items within a booklet. That is, these omit rates remained within the acceptable range for large-scale achievement tests. In summary, the low omit rates observed across entire forms are consistent with tests that are not speeded.

More detailed approaches to check for speededness include examining the relationships of test scores between test books that measure similar content and student performance on individual test books. Beyond omit rates, a test that is not speeded should show empirical relationships among and across all test books. In other words, students performing well on multiple-choice items in Book 1 would also be expected to perform well on multiple-choice items in Book 2. In the presence of speededness, scores on books measuring similar content would exhibit low correlations.

Correlation analysis was conducted to compare the relationship between student performance on each book with student performance on the remaining books. The results are presented in Table

10A–10B and it can be seen that the correlation coefficients were sufficiently high and consistent across books for both ELA and Mathematics Tests. The patterns in both ELA and mathematics reflect what would be expected for tests that do not exhibit speededness.

Table 10A. ELA Correlations (across Books)

Grade	Correlation with Other Books		
	Book 1	Book 2	Book 3
3	.72	.75	.70
4	.70	.77	.70
5	.73	.76	.71
6	.74	.79	.72
7	.74	.79	.73
8	.73	.78	.71

Table 10B. Mathematics Correlations (across Books)

Grade	Correlation with Other Books		
	Book 1	Book 2	Book 3
3	.85	.85	.82
4	.88	.90	.88
5	.87	.87	.86
6	.86	.88	.87
7	.86	.88	.88
8	.83	.84	.84

Next, correlation analysis of student performance was performed on items from each of the books (i.e., book by book) containing similar item types. The results of this analysis set, as presented in Table 11A–11B, were similar to those from the previous analysis: the correlations between individual books were positive and moderately strong. Also, no strong evidence of speededness was observed for ELA Books 2 and 3.

Table 11A. ELA Correlations (Book by Book)

Grade	Book	Book 1	Book 2
3	2	.69	
	3	.62	.68
4	2	.69	
	3	.60	.70
5	2	.70	
	3	.65	.70
6	2	.72	
	3	.65	.73
7	2	.72	
	3	.66	.73
8	2	.72	
	3	.63	.71

Table 11B. Mathematics Correlations (Book by Book)

Grade	Book	Book 1	Book 2
3	2	.83	
	3	.78	.78
4	2	.86	
	3	.84	.86
5	2	.83	
	3	.83	.83
6	2	.83	
	3	.82	.84
7	2	.83	
	3	.82	.85
8	2	.79	
	3	.79	.80

Differential Item Functioning

Classical differential item functioning (DIF) analyses are statistical methods for identifying items that are estimated to have functioned differently for one group (i.e., the “focal” group) as compared with another group (i.e., the “reference” group). In other words, DIF analysis only flags items which may later be judged by experts to exhibit DIF, rather than directly detecting bias. First, the psychometric phenomenon of DIF was extensively investigated and experts’ judgments of bias collected when items were field-tested, which reduced the likelihood of including any differentially functioning items on the operational forms for 2014. Turning to the analysis of the 2014 operational data, as discussed in the “Detection of Bias” subsection of Section 3 above, items flagged for DIF do not necessarily indicate item bias. For example, DIF may be attributed to true group differences on the content measured by the item or Type I error, which refers to statistically flagging items that have no true DIF. Operational items flagged for

DIF are given additional scrutiny by content specialists, above and beyond the existing rounds of reviews by New York State educators, and those content specialists make the final judgment of whether an item is biased for or against the focal group.

DIF was evaluated using two methods, both of which involve checks on statistical and practical significance. First, the Mantel-Haenszel (MH) method is employed for MC items. This non-parametric DIF method partitions the sample of examinees into categories based on total raw test scores. It then compares the log-odds ratio of keyed responses for the focal and reference groups. In terms of statistical significance, the Mantel-Haenszel method has a critical value of 6.63 (degrees of freedom = 1 for MC items; $\alpha = .01$) and as far as practical significance is concerned, it is compared to its corresponding delta-value. Delta-values are a commonly used metric in testing that indicates the magnitude of DIF. Typically, delta-values above 1.50 are considered indicative of moderate DIF that should be examined more closely (Zwick, Donoghue, and Grima, 1993). Second, the standardized mean difference (SMD) was computed for CR items. The SMD statistic (Dorans, Schmitt, and Bleistein, 1992) compares the mean scores of reference and focal groups, after adjusting for proficiency differences. The SMD was also evaluated for statistical significance and, in terms of practical significance, a moderate amount of DIF, for or against the focal group, is represented by an SMD with an absolute value between 0.10 and 0.19, inclusive; a large amount of DIF is represented by an SMD with an absolute value of 0.20 or greater.

Classical DIF analyses were conducted on subgroups of the Needs/Resource Capacity Category (focal group: High Needs; reference group: Low Needs), gender (focal group: Female; reference group: Male), ethnicity (focal groups: Black, Hispanic, and Asian; reference group: White), and English language learners (focal group: English language learners; reference group: Non-English language learners). The DIF analyses were conducted using all cases from the clean data sets. Table 12 and Table 13 show the numbers of cases for the subgroups for ELA and mathematics, respectively.

Table 12. ELA Classical DIF Sample N-Counts

Grade	Ethnicity			Gender		Needs/Resource Capacity Category	
	Black/African American	Hispanic/Latino	White	Female	Male	High	Low
3	35,369	50,691	90,032	97,653	100,835	104,007	75,507
4	35,717	49,427	96,167	100,639	103,161	101,875	78,486
5	35,520	47,929	94,624	98,168	101,259	99,786	80,670
6	37,103	48,052	94,470	99,117	101,323	97,696	79,305
7	36,896	46,537	94,160	97,174	100,429	99,227	81,110
8	38,059	47,207	96,780	99,644	103,714	100,624	81,659

Table 13. Mathematics Classical DIF Sample N-Counts

Grade	Ethnicity			Gender		Needs/Resource Capacity Category	
	Black/African American	Hispanic/Latino	White	Female	Male	High	Low
3	35,570	51,852	89,925	98,086	102,192	105,784	75,402
4	35,799	50,242	95,770	100,866	103,902	103,292	78,016
5	35,284	48,312	93,249	97,562	101,008	100,892	78,472
6	36,931	48,743	92,322	98,050	101,080	98,741	76,939
7	36,433	46,771	90,092	94,663	98,680	99,563	76,543
8	33,981	42,623	68,001	76,750	82,554	86,984	52,526

Table 14 (ELA) and Table 15 (mathematics) present the number of items flagged for DIF by either of the classical methods described earlier. A detailed list of items flagged by either one or both of these classical DIF methods, including DIF direction and associated DIF statistics, is presented in Appendix N.

Table 14. ELA Items Flagged for DIF

Grade	Flagged Items
3	0
4	5
5	5
6	5
7	5
8	8

Table 15. Mathematics Items Flagged for DIF

Grade	Flagged Items
3	2
4	3
5	2
6	2
7	4
8	3

As discussed previously (Section 3, Validity), items showing statistically significant DIF (flagged as described above for MH statistics on MC items and SMD statistics for CR items) do not necessarily pose bias. The items flagged with DIF were examined by the content experts again, and no sign of potential bias was found. In other words, based on combinations of statistical and content evaluations, none of the items on the 3–8 tests showed bias.

Section 6: IRT Calibration and Equating

IRT Models and Rationale for Use

IRT allows for comparisons among items and scale scores, even those from different test forms, by using a common scale for all items and examinees (i.e., as if there were a hypothetical test that contained items from all forms). The three-parameter logistic (3PL) model (Lord and Novick, 1968; Lord, 1980) was used to analyze item responses on the MC items. For analysis of the CR items, the two-parameter partial credit (2PPC) model (Muraki, 1992; Yen, 1993) was used.

IRT is a statistical methodology that takes into account the fact that not all test items are alike and that all items do not provide the same amount of information in determining how much a student knows or can do. Computer programs that implement IRT models use actual student data to estimate the characteristics of the items on a test, called “parameters.” The parameter estimation process is called “item calibration.”

IRT models typically vary according to the number of parameters estimated. For the New York State tests, three parameters are estimated: the discrimination parameter, the difficulty parameter(s), and, for MC items, the guessing parameter. The discrimination parameter is an index of how well an item differentiates between high-performing and low-performing students. An item that cannot be answered correctly by low-performing students, but can be answered correctly by high-performing students, will have a high-discrimination value. The difficulty parameter is an index of how easy or difficult an item is. The higher the difficulty parameter is, the harder the item is. The guessing parameter is the probability that a student with very low proficiency will answer the item correctly.

Because the characteristics of MC and CR items are different, two IRT models were used in item calibration. The three-parameter logistic (3PL) model was used in the analysis of MC items. In this model, the probability that a student with proficiency θ responds correctly to item i is

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + \exp[-1.7a_i(\theta - b_i)]},$$

where

a_i is the item discrimination, b_i is the item difficulty, and c_i is the probability of a correct response from a very low-scoring student.

For analysis of the CR items, the 2PPC model was used. The 2PPC model is a special case of Bock's (1972) nominal model. Bock's model states that the probability of an examinee with proficiency θ having a score $(k - 1)$ at the k th level of the j th item is:

$$P_{jk}(\theta) = P(x_j = k - 1 | \theta) = \frac{\exp Z_{jk}}{\sum_{i=1}^{m_j} \exp Z_{ji}}, \quad k = 1 \dots m_j,$$

where

$$Z_{jk} = A_{jk}\theta + C_{jk},$$

and

k is the item response category ($k = 1, 2, \dots, m_j$).

The m_j denotes the number of score levels for the j th item, and typically the highest score level is assigned $(m_j - 1)$ score points. For the special case of the 2PPC model used here, the following constraints were used:

$$A_{jk} = \alpha_j(k - 1),$$

and

$$C_{jk} = -\sum_{i=0}^{k-1} \gamma_{ji},$$

where

$$\gamma_{j0} = 0,$$

and

α_j and γ_{ji} are the free parameters to be estimated from the data.

Each item has $(m_j - 1)$ independent γ_{ji} parameters and one α_j parameter; a total of m_j parameters are estimated for each item.

Calibration Sample

The cleaned sample data were used for calibration and equating of NYSTP 2014 Grades 3–8 Common Core ELA and Mathematics Tests. It should be noted that the sample sizes were adequate as the calibration and equating were performed using nearly all (96–99%, depending on grade level) of the New York State public and non-public school student population data in each tested grade. As shown in Table 16A–16C and Table 17A–17C, the 2014 operational test samples were generally comparable to 2013 populations in terms of NRC, student race and ethnicity, proportions of English language learners, proportions of students with disabilities, and proportions of students using testing accommodations.

The only subject and grade for which the sample characteristics differed by more than 3.1% for any of those key characteristics was the Grade 8 mathematics sample, and the differences are likely due to the differences in testing requirements that existed between the 2013 and 2014 administrations, in light of the United States Department of Education's granting of a double

testing waiver for accelerated Grades 7 and 8 mathematics students. Those students—who will take the Integrated Algebra Regents Examination—were historically required to also take the Grade 8 mathematics test, but given the waiver, those students’ schools had the discretion to not require such double testing. For more information, see: <http://www.p12.nysed.gov/irs/memos/2013/Submission-of-Request-to-the-USDE-for-a-Waiver-of-the-Requirement-that-Students-in-Grades-7-and-8.pdf>.

Table 16A. ELA Grades 3 and 4 Demographic Statistics

Demographic Category		Grade 3		Grade 4	
		2013 Population	2014 Sample	2013 Population	2014 Sample
Gender	Female	48.84	49.20	49.17	49.38
	Male	51.16	50.80	50.83	50.62
Ethnicity	Asian	8.58	8.63	8.15	8.81
	Black	17.70	17.82	17.36	17.53
	Hispanic	23.83	25.54	23.14	24.25
	American Indian	0.54	0.58	0.53	0.55
	Multiracial	1.38	1.81	1.19	1.45
	Pacific Islander	0.22	0.26	0.22	0.22
	White	47.75	45.36	49.41	47.19
NRC	New York	35.30	35.76	32.71	34.01
	Big 4 Cities	3.90	3.90	3.69	3.80
	Urban/Suburban	7.49	7.58	7.22	7.12
	Rural	5.42	5.21	5.16	5.10
	Average Needs	27.62	25.91	27.11	25.79
	Low Needs	13.39	12.17	13.16	12.76
	Charter	3.63	4.19	2.68	3.47
	Non-Public	3.24	5.27	8.26	7.94
SWD	No	86.18	85.54	85.83	85.08
	Yes	13.82	14.46	14.17	14.92
SUA	No	86.27	87.96	87.62	87.04
	Yes	13.73	12.04	12.38	12.96
ELL	No	91.38	92.15	93.15	92.02
	Yes	8.62	7.85	6.85	7.98

Table 16B. ELA Grades 5 and 6 Demographic Statistics

Demographic Category		Grade 5		Grade 6	
		2013 Population	2014 Sample	2013 Population	2014 Sample
Gender	Female	49.12	49.23	49.12	49.45
	Male	50.88	50.77	50.88	50.55
Ethnicity	Asian	8.24	8.63	7.90	8.46
	Black	18.03	17.81	18.43	18.51
	Hispanic	22.94	24.03	22.80	23.97
	American Indian	0.55	0.57	0.51	0.55
	Multiracial	1.05	1.27	0.96	1.17
	Pacific Islander	0.17	0.23	0.20	0.20
	White	49.02	47.45	49.20	47.13
NRC	New York	33.83	33.74	31.87	32.75
	Big 4 Cities	3.83	3.84	3.68	3.72
	Urban/Suburban	7.38	7.27	6.99	7.08
	Rural	5.48	5.25	5.35	5.24
	Average Needs	28.65	27.13	28.00	26.38
	Low Needs	14.08	13.37	13.61	13.23
	Charter	3.45	4.08	3.07	4.08
	Non-Public	3.30	5.33	7.41	7.53
SWD	No	84.67	84.25	85.34	84.70
	Yes	15.33	15.75	14.66	15.30
SUA	No	85.96	87.21	86.73	87.08
	Yes	14.04	12.79	13.27	12.92
ELL	No	92.92	93.38	94.95	93.37
	Yes	7.08	6.62	5.05	6.63

Table 16C. ELA Grades 7 and 8 Demographic Statistics

Demographic Category		Grade 7		Grade 8	
		2013 Population	2014 Sample	2013 Population	2014 Sample
Gender	Female	48.80	49.18	48.99	49.00
	Male	51.20	50.82	51.01	51.00
Ethnicity	Asian	8.31	8.35	7.87	8.79
	Black	18.30	18.67	18.49	18.72
	Hispanic	22.02	23.55	21.94	23.21
	American Indian	0.50	0.55	0.49	0.53
	Multiracial	0.90	1.03	0.86	0.95
	Pacific Islander	0.19	0.20	0.18	0.21
	White	49.80	47.65	50.17	47.59
NRC	New York	33.40	33.74	31.81	33.89
	Big 4 Cities	3.69	3.79	3.44	3.54
	Urban/Suburban	7.20	7.18	6.80	6.80
	Rural	5.64	5.57	5.43	5.30
	Average Needs	29.09	26.90	27.96	26.10
	Low Needs	15.36	14.20	14.69	14.09
	Charter	2.62	3.26	2.07	2.49
	Non-Public	3.01	5.37	7.78	7.78
SWD	No	85.02	84.63	85.82	85.58
	Yes	14.98	15.37	14.18	14.42
SUA	No	87.35	89.19	88.58	89.26
	Yes	12.65	10.81	11.42	10.74
ELL	No	95.08	94.50	94.88	94.74
	Yes	4.92	5.50	5.12	5.26

Table 17A. Mathematics Grades 3 and 4 Demographic Statistics

Demographic Category		Grade 3		Grade 4	
		2013 Population	2014 Sample	2013 Population	2014 Sample
Gender	Female	48.79	48.97	49.18	49.26
	Male	51.21	51.03	50.82	50.74
Ethnicity	Asian	8.77	8.82	8.32	8.99
	Black	17.63	17.76	17.28	17.48
	Hispanic	24.04	25.89	23.32	24.54
	American Indian	0.54	0.58	0.53	0.55
	Multiracial	1.37	1.79	1.18	1.45
	Pacific Islander	0.22	0.26	0.22	0.23
	White	47.43	44.90	49.15	46.77
NRC	New York	35.58	36.18	32.95	34.43
	Big 4 Cities	3.90	3.96	3.73	3.86
	Urban/Suburban	7.51	7.60	7.23	7.16
	Rural	5.39	5.14	5.12	5.05
	Average Needs	27.43	25.57	26.90	25.48
	Low Needs	13.34	12.12	13.08	12.66
	Charter	3.64	4.16	2.70	3.45
	Non-Public	3.20	5.28	8.29	7.92
SWD	No	85.90	85.39	85.53	85.03
	Yes	14.10	14.61	14.47	14.97
SUA	No	83.57	83.40	85.66	82.55
	Yes	16.43	16.60	14.34	17.45
ELL	No	90.46	91.15	92.18	91.05
	Yes	9.54	8.85	7.82	8.95

Table 17B. Mathematics Grades 5 and 6 Demographic Statistics

Demographic Category		Grade 5		Grade 6	
		2013 Population	2014 Sample	2013 Population	2014 Sample
Gender	Female	49.14	49.13	49.18	49.24
	Male	50.86	50.87	50.82	50.76
Ethnicity	Asian	8.39	8.87	8.08	8.71
	Black	17.98	17.77	18.35	18.55
	Hispanic	23.22	24.33	23.07	24.48
	American Indian	0.55	0.57	0.51	0.54
	Multiracial	1.05	1.28	0.95	1.15
	Pacific Islander	0.17	0.23	0.20	0.22
	White	48.63	46.96	48.84	46.36
NRC	New York	34.11	34.45	32.12	33.57
	Big 4 Cities	3.86	3.89	3.72	3.79
	Urban/Suburban	7.40	7.33	6.93	7.11
	Rural	5.45	5.20	5.36	5.16
	Average Needs	28.45	26.29	27.77	25.69
	Low Needs	14.00	13.27	13.50	12.98
	Charter	3.49	4.10	3.14	4.10
	Non-Public	3.22	5.47	7.44	7.60
SWD	No	84.36	84.27	85.06	84.77
	Yes	15.64	15.73	14.94	15.23
SUA	No	83.93	82.28	84.84	83.22
	Yes	16.07	17.72	15.16	16.78
ELL	No	91.90	92.39	93.87	92.34
	Yes	8.10	7.61	6.13	7.66

Table 17C. Mathematics Grades 7 and 8 Demographic Statistics

Demographic Category		Grade 7		Grade 8	
		2013 Population	2014 Sample	2013 Population	2014 Sample
Gender	Female	48.83	48.96	49.03	48.18
	Male	51.17	51.04	50.97	51.82
Ethnicity	Asian	8.47	8.62	8.04	7.56
	Black	18.23	18.84	18.44	21.33
	Hispanic	22.32	24.19	22.22	26.76
	American Indian	0.50	0.55	0.48	0.56
	Multiracial	0.89	1.00	0.85	0.91
	Pacific Islander	0.19	0.20	0.19	0.20
	White	49.41	46.60	49.78	42.69
NRC	New York	33.79	34.98	32.20	37.18
	Big 4 Cities	3.72	3.88	3.49	4.28
	Urban/Suburban	7.22	7.24	6.82	7.61
	Rural	5.62	5.45	5.45	5.57
	Average Needs	28.78	25.94	27.61	22.98
	Low Needs	15.21	13.69	14.54	10.02
	Charter	2.58	3.32	2.05	3.05
	Non-Public	3.07	5.50	7.83	9.30
SWD	No	84.86	84.71	85.69	82.85
	Yes	15.14	15.29	14.31	17.15
SUA	No	87.41	87.35	87.91	83.98
	Yes	12.59	12.65	12.09	16.02
ELL	No	94.06	93.37	93.88	92.19
	Yes	5.94	6.63	6.12	7.81

Calibration Process

The item parameters were estimated using MULTILOG software (Thissen, 1991). MC and CR items were calibrated simultaneously using marginal maximum likelihood procedures.

The calibration of NYSTP 2014 Grades 3–8 Common Core ELA and Mathematics Tests did not exhibit any test-level issues. The estimated parameters were on the original theta scale, and all the items were well within the prescribed parameter ranges. For both the Grades 3–8 Common Core ELA and Mathematics Tests, all calibration estimation results were reasonable. The summaries of the calibration results are presented in Table 18 for ELA and Table 19 for mathematics. Additional detail, including individual item parameter estimates, may be found in Appendix O, in Tables O13–O24. The parameter estimates are expressed on the theta metric and are defined below:

- for MC items:
 - a -parameter is a discrimination parameter;
 - b -parameter is a difficulty parameter; and
 - c -parameter is a guessing parameter;
- for CR items:
 - α is a discrimination parameter; and
 - $step$ is a difficulty parameter for category m_j .

As described in Section 6 “IRT Calibration and Equating,” above in subsection “IRT Models and Rationale for Use,” m_j denotes the number of score levels for the j th item, and typically the highest score level is assigned ($m_j - 1$) score points. Note that for the 2PPC model there are $m_j - 1$ independent steps and one alpha, for a total of m_j independent parameters estimated for each item, while there is one a -parameter and one b -parameter per item in the 3PL model.

Table 18. ELA Calibration Results

Grade	Item-level			Student-Level		
	Largest a-Parameter	Range of b- / Step Parameters		N-Count	Theta Est.*	
					Mean	SD
3	1.443	-2.918	2.573	198,488	0.01	0.93
4	0.977	-3.018	3.503	203,800	0.00	0.94
5	1.223	-2.665	2.674	199,427	0.01	0.94
6	1.426	-2.780	2.034	200,440	0.01	0.94
7	1.531	-3.141	2.246	197,603	0.01	0.94
8	1.154	-3.177	1.686	203,358	0.01	0.94

*: Maximum *a posteriori* (MAP) theta estimates.

Table 19. Mathematics Calibration Results

Grade	Item-level			Student-Level		
	Largest a-Parameter	Range of b- / Step Parameters		N-Count	Theta Est.*	
					Mean	SD
3	1.744	-2.613	2.441	200,278	0.00	0.94
4	1.590	-2.443	1.233	204,768	0.01	0.95
5	1.730	-2.384	2.259	198,570	0.02	0.95
6	1.925	-2.204	1.933	199,130	0.02	0.95
7	2.652	-1.951	2.192	193,343	0.02	0.95
8	1.980	-2.119	2.119	159,304	0.02	0.94

*: Maximum *a posteriori* (MAP) theta estimates.

Item-Model Fit

Item fit statistics provide evidence of the appropriateness of using an item in the 3PL or 2PPC model. The Q_I procedure described by Yen (1981) was used to measure fit to the three-parameter model. Students are rank-ordered on the basis of $\hat{\theta}$ values and sorted into ten cells with 10% of the sample in each cell. For each item, the number of students in cell k who answered item i , N_{ik} , and the number of students in that cell who answered item i correctly, R_{ik} , were determined. The observed proportion in cell k passing item i , O_{ik} , is R_{ik}/N_{ik} . The fit index for item i is:

$$Q_{Ii} = \sum_{k=1}^{10} \frac{N_{ik} (O_{ik} - E_{ik})^2}{E_{ik} (1 - E_{ik})},$$

with:

$$E_{ik} = \frac{1}{N_{ik}} \sum_{j \in \text{cell } k}^{N_{ik}} P_i(\hat{\theta}_j).$$

A modification of this procedure was used to measure fit to the 2PPC model. For the 2PPC model, Q_{Ij} was assumed to have an approximate chi-square distribution with the following degrees of freedom (df):

$$df = I(m_j - 1) - m_j,$$

where I is the total number of cells (usually 10) and m_j is the possible number of score levels for item j .

To adjust for differences in degrees of freedom among items, Q_I was transformed to Z_{Q_I} where:

$$Z_{Q_I} = (Q_I - df) / (2df)^{1/2}.$$

The value of Z increases with sample size, when all else is equal. To use this standardized statistic to flag items for potential poor fit, it has been a common practice to vary the critical value for Z as a function of sample size. For the tests that have large calibration sample sizes, the criterion $Z_{Q_I} \text{Crit}$ was used to flag items and was calculated using the expression

$$Z_{Q_I} \text{Crit} = \left(\frac{N}{1500} \right) * 4,$$

where N is the calibration sample size.

To compute the Q_I and related statistics, a stratified sampling procedure was implemented in a way that a representative sample with the size of approximately 700,000 students was drawn at each grade level. Items were considered to have poor fit if the value of the obtained Z_{Q_I} was

greater than the value of Z_{Q_1} critical. If the obtained Z_{Q_1} was less than Z_{Q_1} critical, the items were rated as having acceptable fit. The fact that the majority of the items in the NYSTP 2014 Grades 3–8 Common Core ELA and Mathematics Tests demonstrated good model fit further supports the use of the chosen models. Item fit statistics are presented in Appendix O, in Tables O1–O12.

Local Independence

In using IRT models, one of the assumptions made is that the items are locally independent, that a student's response on one item is not dependent upon his or her response to another item. In other words, when a student's proficiency is accounted for, his or her response to each item is statistically independent.

One way to measure the statistical independence of items within a test is via the Q_3 statistic (Yen, 1984). This statistic was obtained by correlating differences between students' observed and expected responses for pairs of items after taking into account overall test performance. The Q_3 statistic for binary items was computed as

$$d_{ij} \equiv u_{ij} - P_j(\hat{\theta}_i)$$

where $\hat{\theta}_i$ is the estimated trait value (i.e., proficiency) for the i th examinee; u_{ij} is the observed probability for the i th examinee to get the j th item correct and P_j is estimated probability for the i th examinee to get the j th item correct, and

$$Q_{3jj'} = r(d_j, d_{j'}).$$

The generalization to items with multiple response categories uses

$$d_{ij} \equiv x_{ij} - E_{ij},$$

where

$$E_{ij} \equiv E(x|\hat{\theta}_i) = \sum_{k=1}^{m_j} kP_{jk}(\hat{\theta}_i).$$

If a substantial number of items in the test demonstrate local dependence, these items may need to be calibrated separately. All pairs of items with Q_3 values greater than .20 were classified as significant for local dependency. The maximum value for this index is 1.00. When item pairs are flagged by Q_3 , the content of the flagged items is examined to identify possible sources of the local dependence. The primary concern about locally dependent items is that they contribute less psychometric information about examinee proficiency than do locally independent items and therefore inflate score reliability estimates.

The Q_3 statistics were examined for all unique pairs of ELA and mathematics items. Items that were found to be significant in local dependency vary depending on the subject and grade: two pairs of items each were found in ELA Grades 6 and 7; and one pair of items was found in Grade

8. When reviewing the results for mathematics two pairs of items exceeded a correlation of .20 in mathematics Grade 4 and one pair of items each in Grades 3, 5, 6, and 8. The magnitudes of these statistics were not sufficient to warrant further concern or action (with the Q_3 values ranging from .20 to .39 for ELA tests and from .20 to .43 for mathematics tests).

Equating and Scaling

With the new assessments being implemented in 2013, a new scale was established after the data were collected. The purpose of equating was to place the 2014 item parameters and proficiency estimates on the same scale as those in 2013. The following steps constitute the equating process for each subject and grade:

1. Operational items as well as non-scored (i.e., external) anchor items were calibrated in MULTILOG.
2. The 2014 item parameter estimates for all anchor items—both scored and non-scored—enabled the establishment of the equating relationship via a test characteristic curve (TCC) method (Stocking and Lord, 1983; implemented in STUIRT, Kim, & Kolen, 2004) to the 2013 theta scale, using the established 2013 item parameter estimates for those same items. The resulting equating coefficients are displayed in Table 20 and
3. Table 21, and the following parameters were equated using the formulas below:

$$a_i^E = a_i^C / M_1^E ,$$

$$b_i^E = M_1^E \cdot b_i^C + M_2^E , \text{ and}$$

$$d_{ij}^E = d_{ij}^C + \left[\left(a_i^C / M_1^E \right) \right] \cdot M_2^E ,$$

where

M_1^E is defined as the multiplicative adjustment for equating and M_2^E is the additive adjustment for equating. Also note that the superscript “E” denotes equated item parameter estimates, while the superscript “C” denotes calibrated item parameter estimates.

Table 20. ELA Equating Coefficients

Grade	M_1^E	M_2^E
3	1.033	-0.083
4	0.993	-0.015
5	1.026	-0.077
6	0.972	-0.044
7	0.977	-0.126
8	1.015	-0.010

Table 21. Mathematics Equating Coefficients

Grade	M_1^E	M_2^E
3	1.046	0.128
4	1.097	0.119
5	1.064	0.199
6	1.071	0.104
7	1.011	0.093
8	0.998	-0.183

4. A raw-score-to-theta conversion chart was produced using the test characteristic curve (TCC) method (Stocking and Lord, 1983; see the “Scoring Procedure” subsection for more detail) and implemented in POLYEQUATE (Kolen & Cui, 2004). The theta estimates associated with the TCC method ($\hat{\theta}_{TCC}$) must be equated back to the underlying theta scale established in the base year (Spring 2013) and are computed as follows:

$$\theta^E = (M_1^E \cdot \hat{\theta}_{TCC}) + M_2^E$$

5. The TCC method does not produce theta estimates for raw scores below chance level or above the perfect score (highest obtainable raw score). In addition, for the scores at the low and high ends of the scale, some raw scores tended to have large theta estimates (for example, -7.999). Typically the first obtainable theta value on a test corresponds to a very extreme theta value. The following adjustment/interpolation was conducted:

For any equated theta estimates (θ^E) that are outside of the range of -2.5 to 3, at the lower end of the scale, 0.25 was subtracted from the preceding theta value that is within the range; at the higher end of the scale, 0.25 was added to the previous theta value that is within the range, thus resulting in an adjusted theta estimate (θ^A) for those extremes. See the table below for an example at the lower end of the scale. Such an adjustment helps contain the theta scale within a reasonable range and is standard practice in testing.

Raw score	θ^E	θ^A
6	-5.30263	-3.37458
7	-3.66491	-3.12458
8	-3.03055	-2.87458
9	-2.76782	-2.62458
10	-2.37458	-2.37458

- Once theta values were either estimated or interpolated for all raw scores, the raw-score-to-theta relationship was applied to each student, yielding a theta estimate corresponding to his or her raw score.
- The adjusted theta estimates were then scaled using the established scaling coefficients from the base year (Spring 2013) and displayed in Table 22 and Table 23 according to the following formula:

$$ScaleScore = (M_1^S \cdot \theta^A) + M_2^S,$$

where

M_1^S is defined as the multiplicative scaling coefficient, and M_2^S is the additive scaling coefficient. M_1^S and M_2^S are applied to a true score (i.e., the post-equated theta estimate) in order to obtain a scale score.

Table 22. ELA Scaling Coefficients

Grade	M_1^S	M_2^S
3	31.814	301.495
4	32.036	300.762
5	32.016	300.954
6	32.258	300.673
7	31.926	300.801
8	31.627	300.980

Table 23. Mathematics Scaling Coefficients

Grade	M_1^S	M_2^S
3	32.249	299.856
4	32.698	300.176
5	32.220	300.693
6	32.421	300.377
7	31.229	301.144
8	31.868	301.143

8. Scale scores range approximately from 100 to 400 across grades. The lowest and highest observed scale score (LOSS and HOSS, respectively) may vary by grade.
9. A series of anchor set stability checks were performed before finalizing the anchor set for each subject and grade; see the “Anchor Set Evaluation” subsection, which follows this one.
10. For conditional standard error of measurement (CSEM), the scale scores (both estimated and interpolated) were used to compute the information function and CSEM.

Throughout this process, NYSED psychometricians have reviewed, and a senior scientist from HumRRO has independently verified, the results generated by Pearson psychometricians. For future years, the same process will be followed:

- Item parameter estimates will be equated to the base scale established in Spring 2013.
- TCC estimates will be calculated using equated item parameter estimates.
- Adjustments/interpolations will be made to thetas outside of the -2.5–3 range or when there are no estimates provided by POLYEQUATE.
- M_1^S and M_2^S established in the base year (Spring 2013) will be applied to the estimated and interpolated thetas from the TCC approach.
- LOSS and HOSS will be computed accordingly.

Anchor Set Evaluation

In order to determine if each item from the anchor set performs similarly to when it was administered in the prior year, comparisons of individual item characteristic curves (ICCs) and item parameter estimates from the previous and current administrations were made. Initial comparisons included a graphical inspection of the linearity of relationships between equated item parameter estimates from the 2013 and 2014 administrations. These revealed approximately linear relationships as well as similarities in item functions and therefore provided support for the selected post-equating method used herein. Additional analyses of the correlations between equated item parameter estimates also provided evidence of strong linear relationships.

A formal process for validating the anchor set using an objective criterion was used to determine if any items ought to be considered for removal from the anchor set. The equated item parameter estimates were used to calculate a weighted, squared deviation of the current ICC from the previous ICC, across the range of ability (i.e., theta, or θ) and under a hypothetical normal distribution for θ . For a given item i , that quantity called “d squared” is given by

$$d_i^2 = \sum_k \{ [P_{i1}(\theta_k) - P_{i0}(\theta_k)]^2 \cdot g(\theta_k) \},$$

where i indexes anchor items; k indexes quadrature points for θ ; $P_{i1}(\cdot)$ is the probability of a correct response to item i under the current calibration, while $P_{i0}(\cdot)$ is the same quantity under the previous calibration; and $g(\theta_k)$ are weights for the quadrature points.

Historically and as recently as the 2012 operational equating, a fixed criterion on this metric ($d_i^2 \geq 0.05$) has been used for flagging items to be considered for removal from equating. The same approach and criterion were used for the equating of the 2014 operational forms to the 2013 scale score scale. This procedure minimizes the weighted squared differences between the two ICCs for each MC item: one based on 2013 item parameter estimates and the other on 2014 estimates. The differential item performance was evaluated by examining previous and current item parameters. The following steps were taken:

1. Before the iterative procedures start, the initial equating was performed, using all the eligible anchor items as an anchor set, as described in the “Equating and Scaling” section. The initial equating coefficients (M_1^E and M_2^E) were obtained through the Stocking-Lord method.
2. The following process was repeated for at least five iterations or until the largest $d_i^2 < 0.05$ is reached, whichever was greater:
 - a. For each anchor item, d_i^2 was calculated as a weighted sum of the squared deviations between the ICCs based on old (2013) and new (2014) parameter estimates at each quadrature point and assuming a normal theta distribution.
 - b. The item having the largest d_i^2 was identified and removed from the anchor set.
 - c. The equating procedures described in the “Equating and Scaling” subsection were performed with the newly reduced anchor set.
 - d. New raw-score-to-scale-score tables were prepared as described in the “Scoring Procedure” subsection.
3. Select the equating coefficients (M_1^E and M_2^E) associated with the iteration selected in step 2 above.

The items that are implicitly proposed for removal from the anchor set based on the process described above were summarized and evaluated. The only subject where items were proposed

and ultimately approved for removal from the anchor set was mathematics, and one item each was removed from the anchor sets for Grades 5, 6, and 7.

Test Characteristic Curves

Test Characteristic Curves (TCCs) provide an overview of the tests in the IRT scale score metric. The 2014 TCCs were generated using final item parameters for all reporting test items administered in Spring 2014. TCCs are the summation of all the item characteristic curves (ICCs) for items that contribute to the scale score. Conditional standard error of measurement (CSEM) curves graphically show the amount of measurement error at different proficiency levels. The TCCs and CSEM curves are presented in Figure 1 through Figure 24.

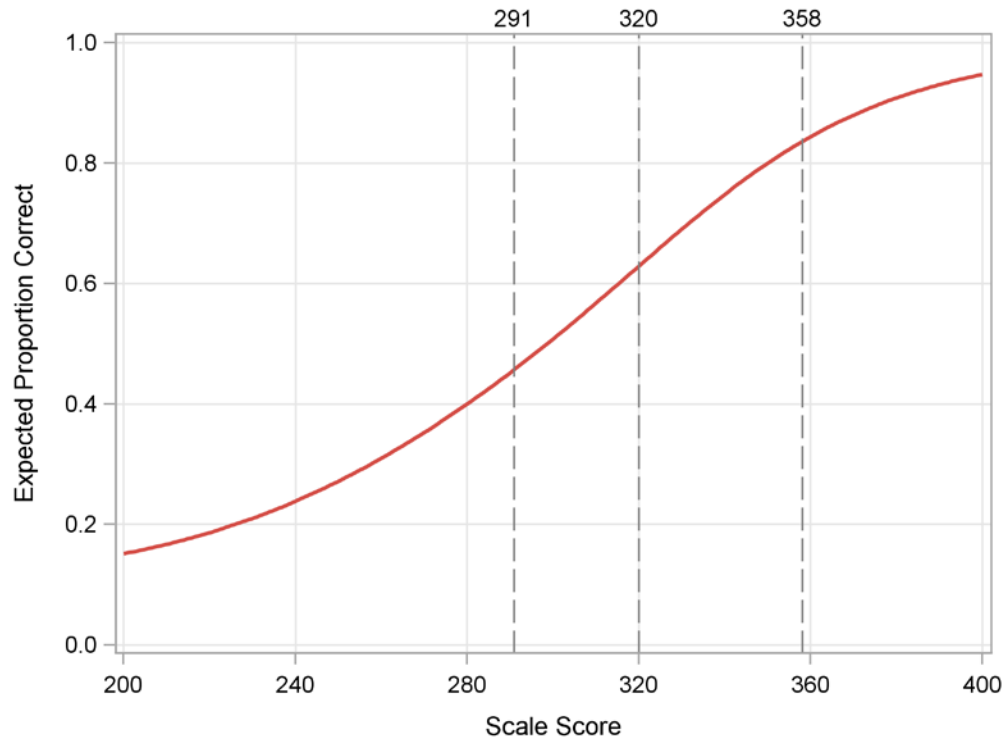


Figure 1. ELA Grade 3 TCC

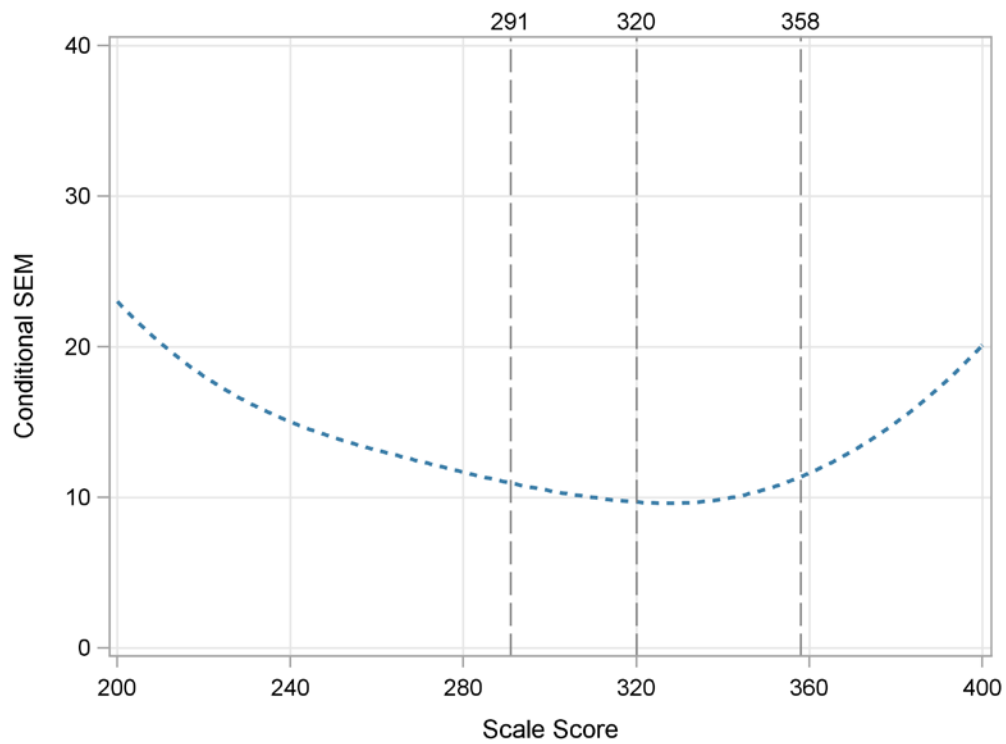


Figure 2. ELA Grade 3 CSEM Curve

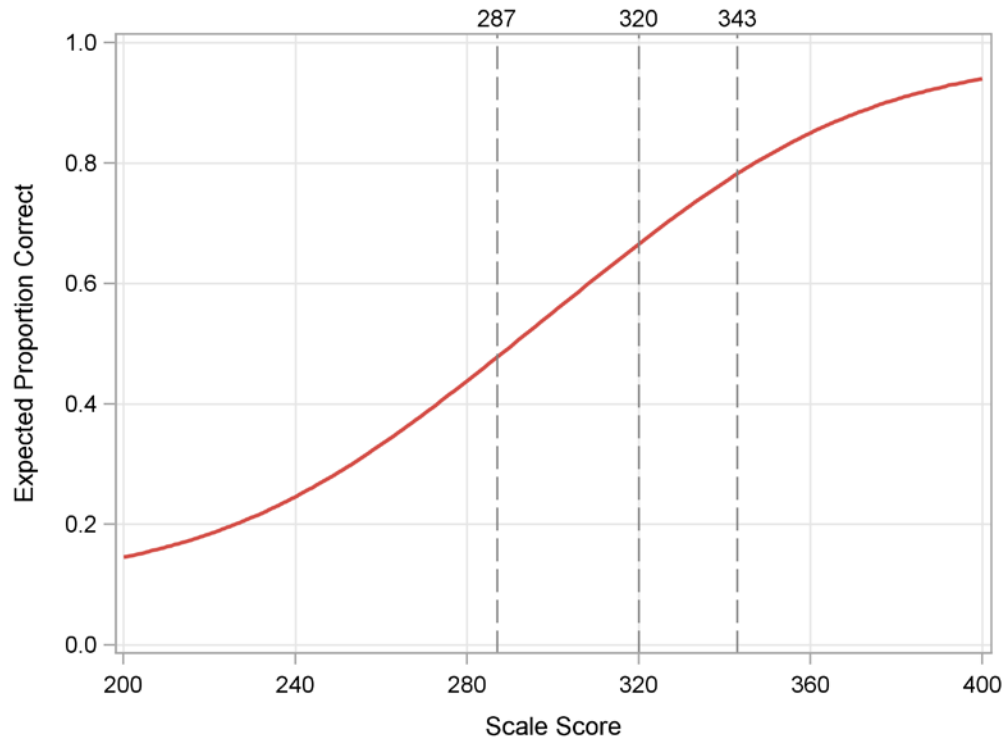


Figure 3. ELA Grade 4 TCC

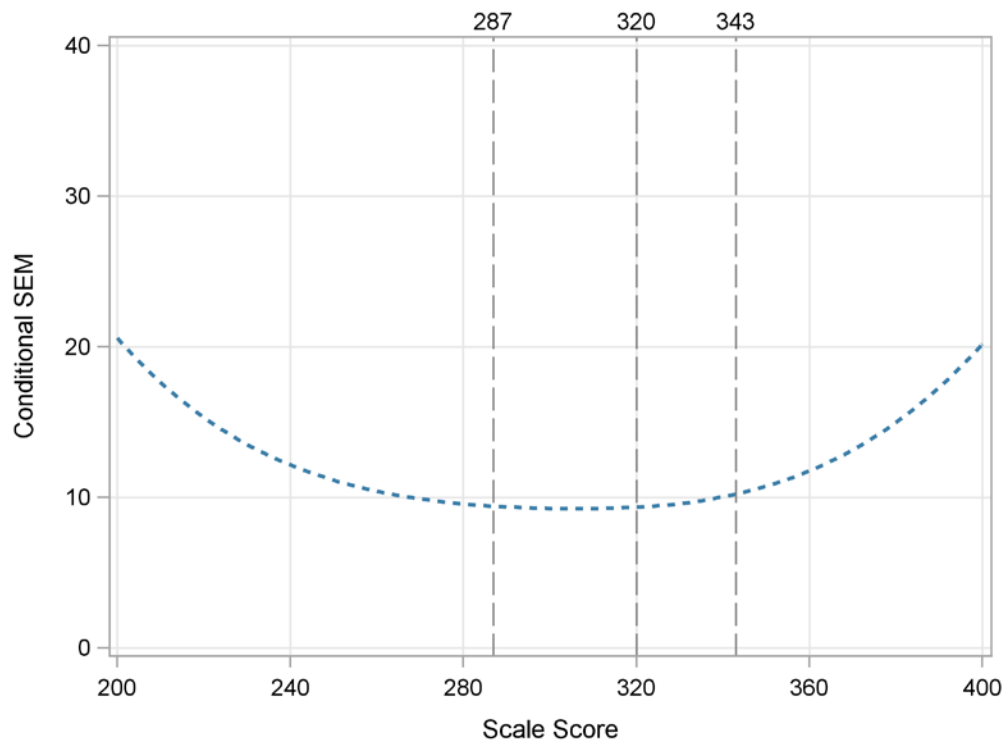


Figure 4. ELA Grade 4 CSEM Curve

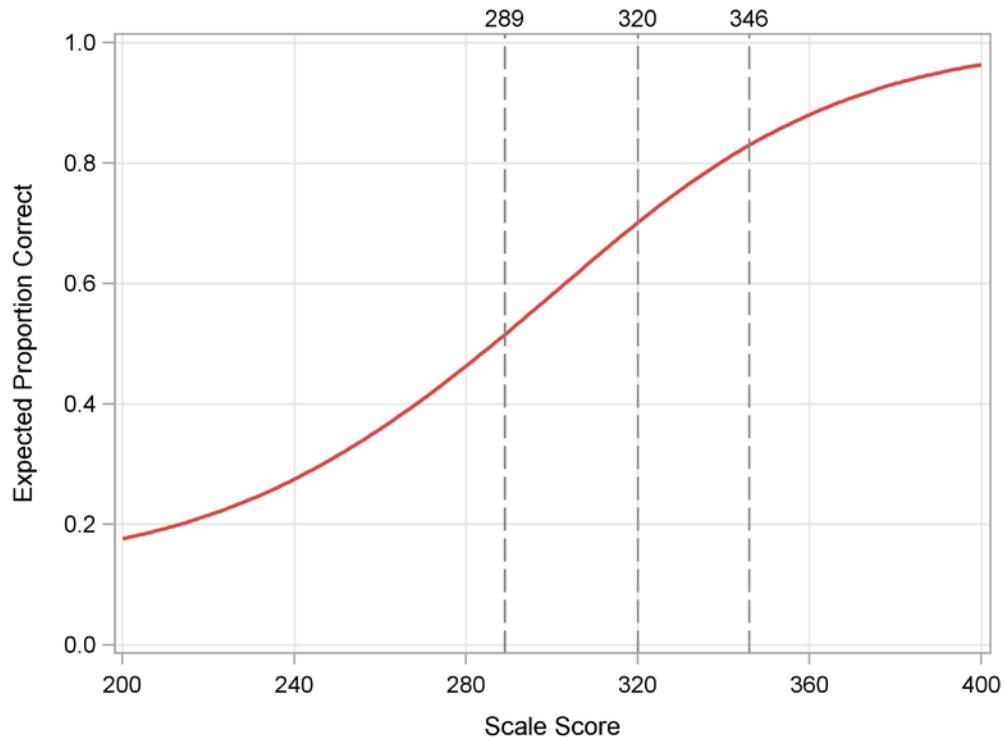


Figure 5. ELA Grade 5 TCC

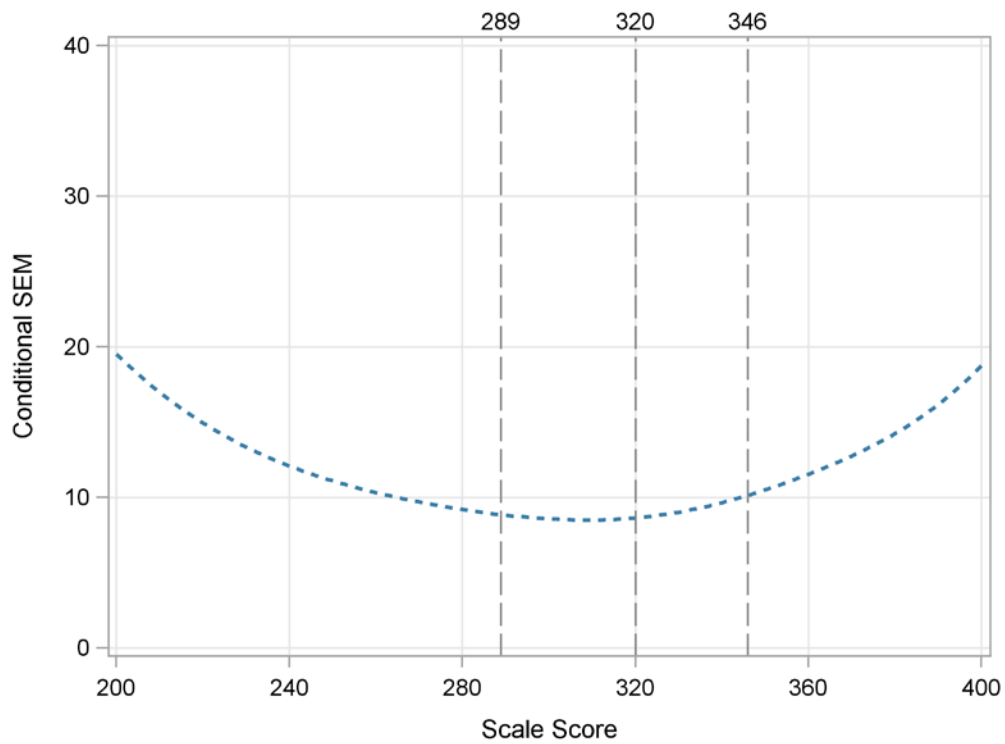


Figure 6. ELA Grade 5 CSEM Curve

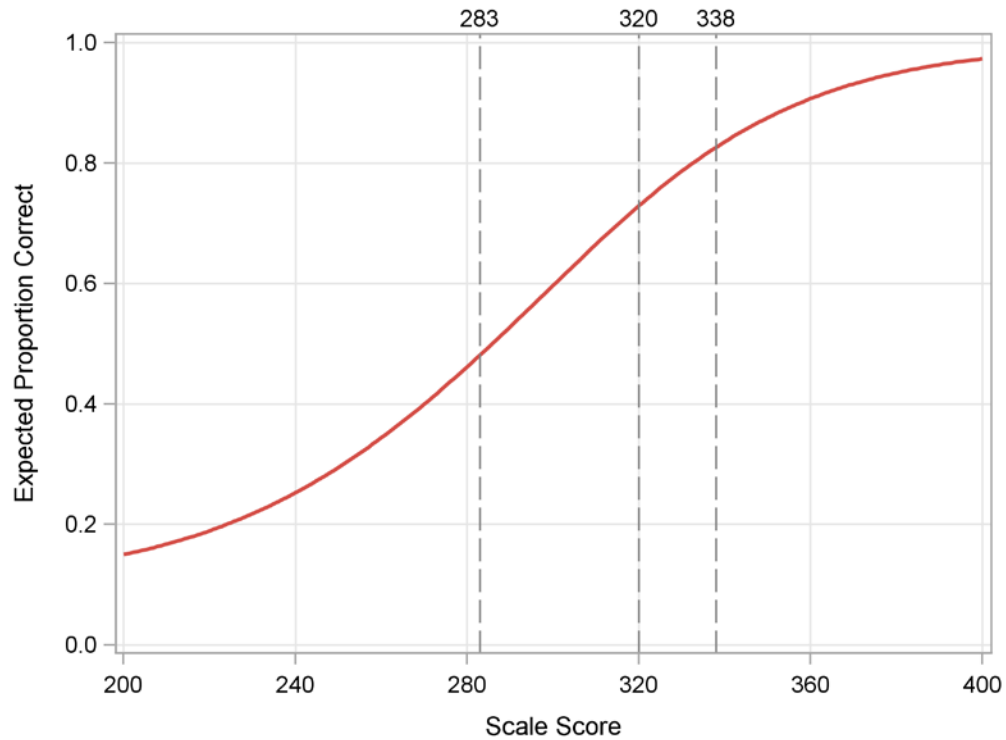


Figure 7. ELA Grade 6 TCC

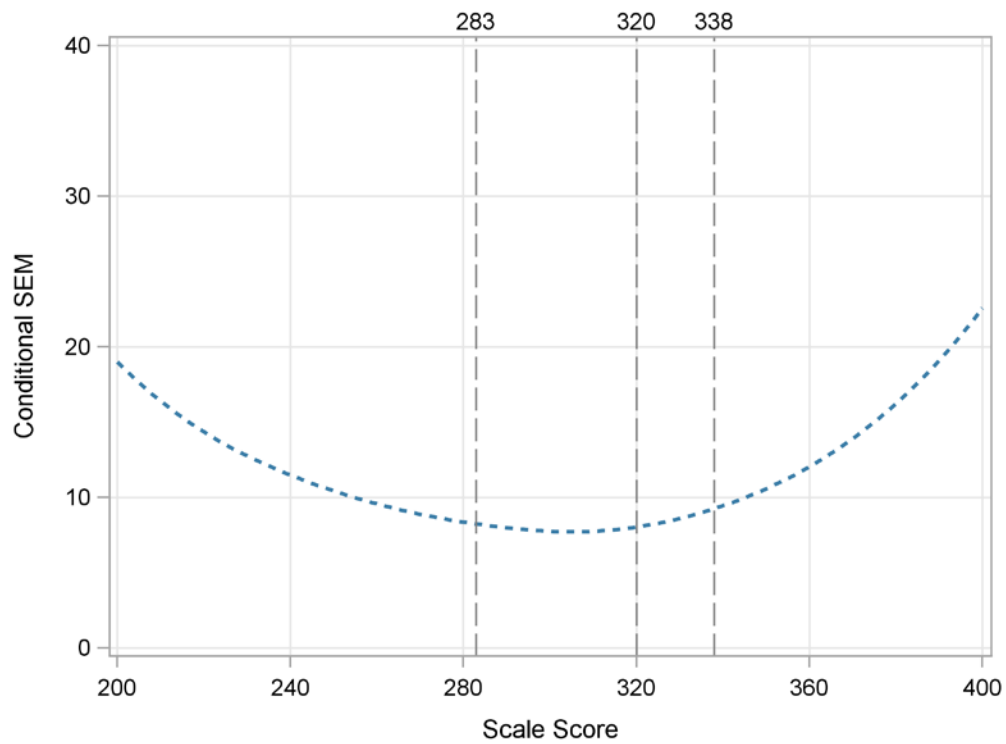


Figure 8. ELA Grade 6 CSEM Curve

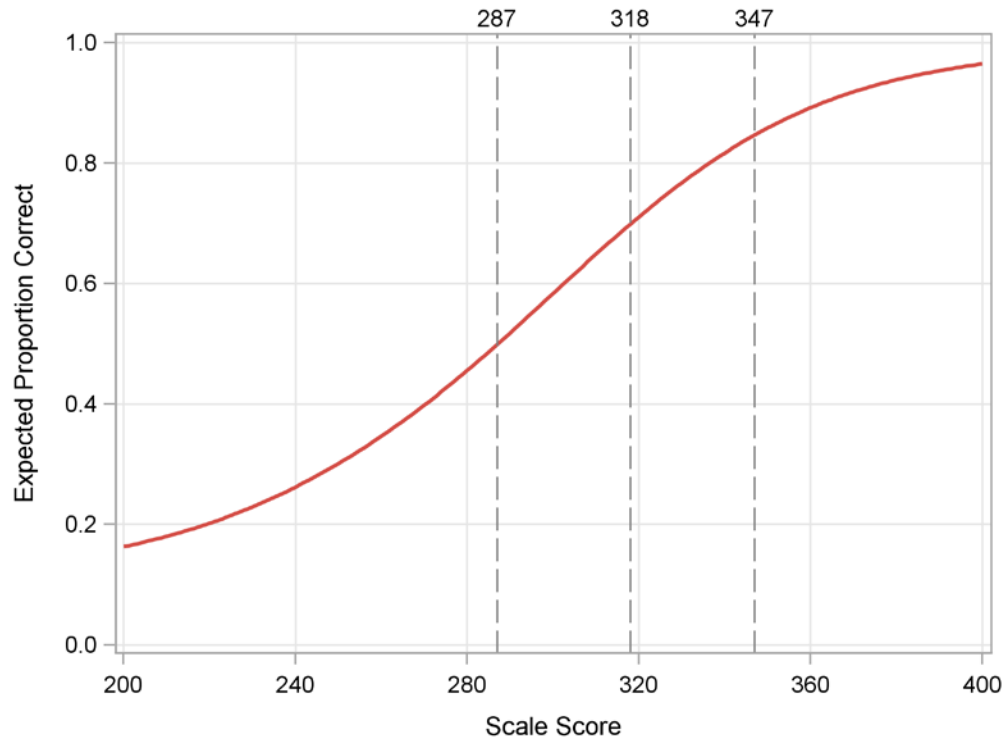


Figure 9. ELA Grade 7 TCC

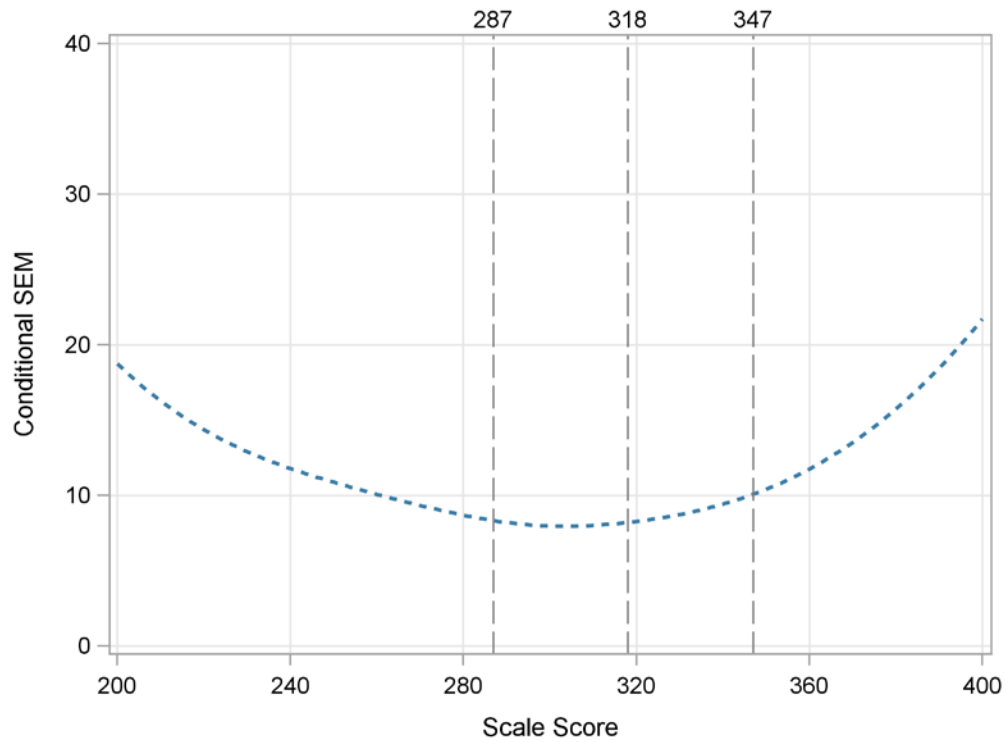


Figure 10. ELA Grade 7 CSEM Curve

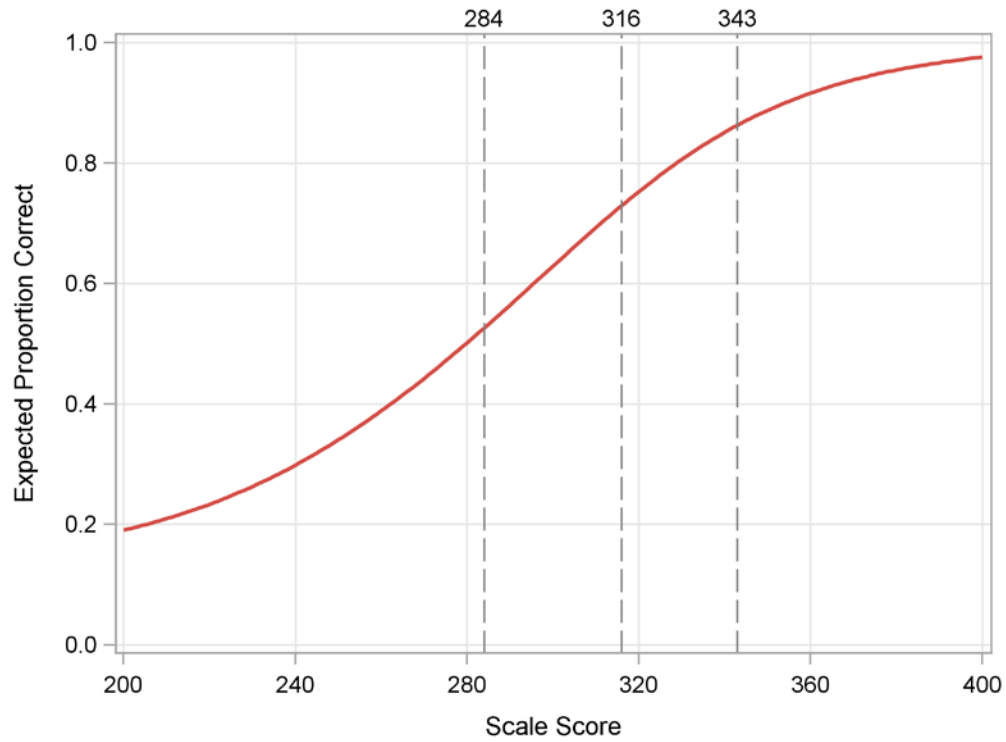


Figure 11. ELA Grade 8 TCC

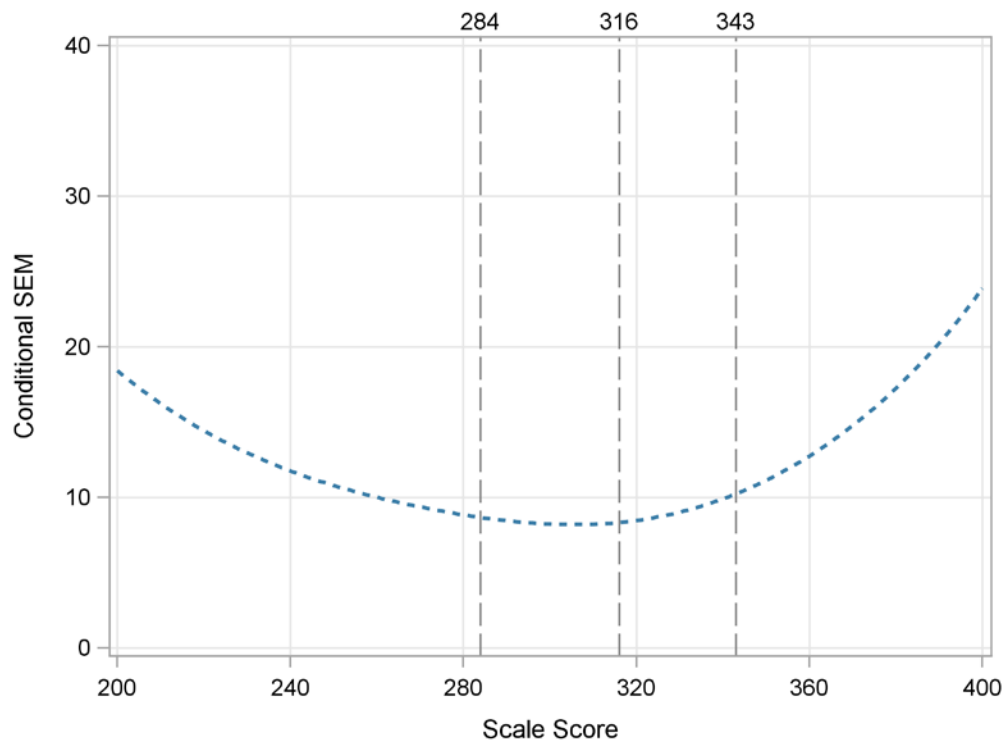


Figure 12. ELA Grade 8 CSEM Curve

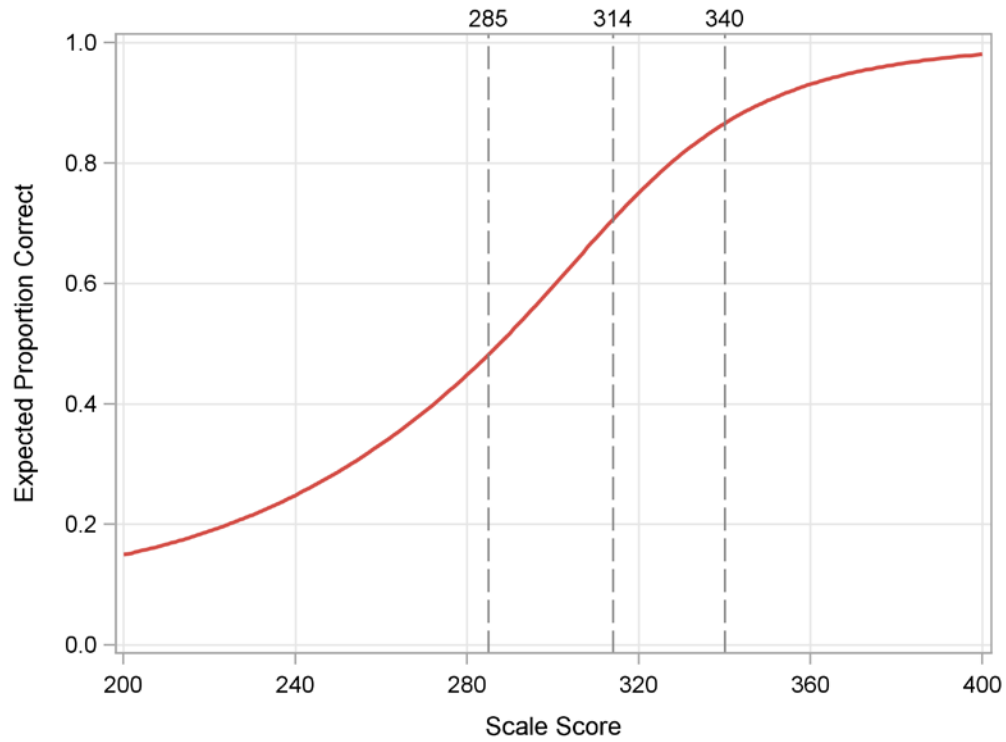


Figure 13. Mathematics Grade 3 TCC

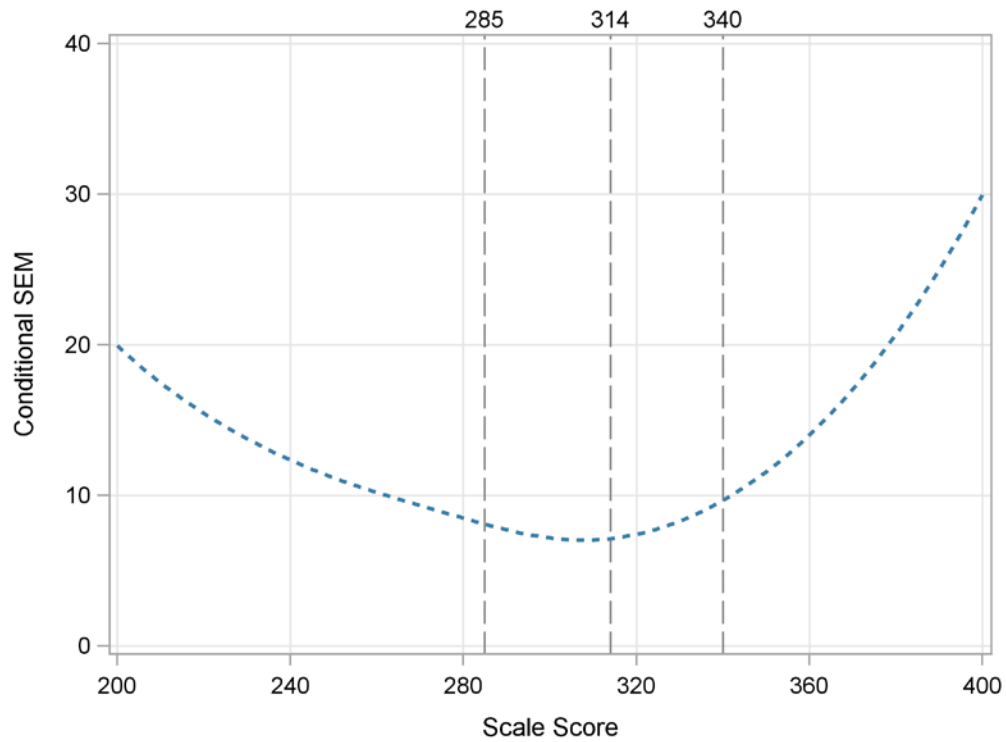


Figure 14. Mathematics Grade 3 CSEM Curve

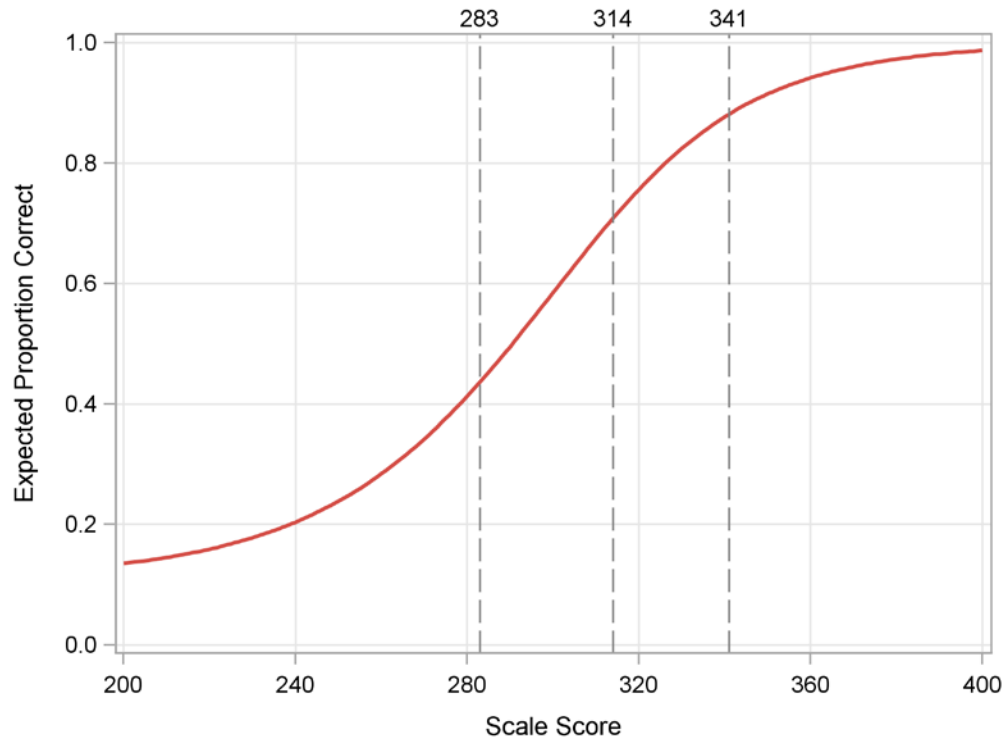


Figure 15. Mathematics Grade 4 TCC

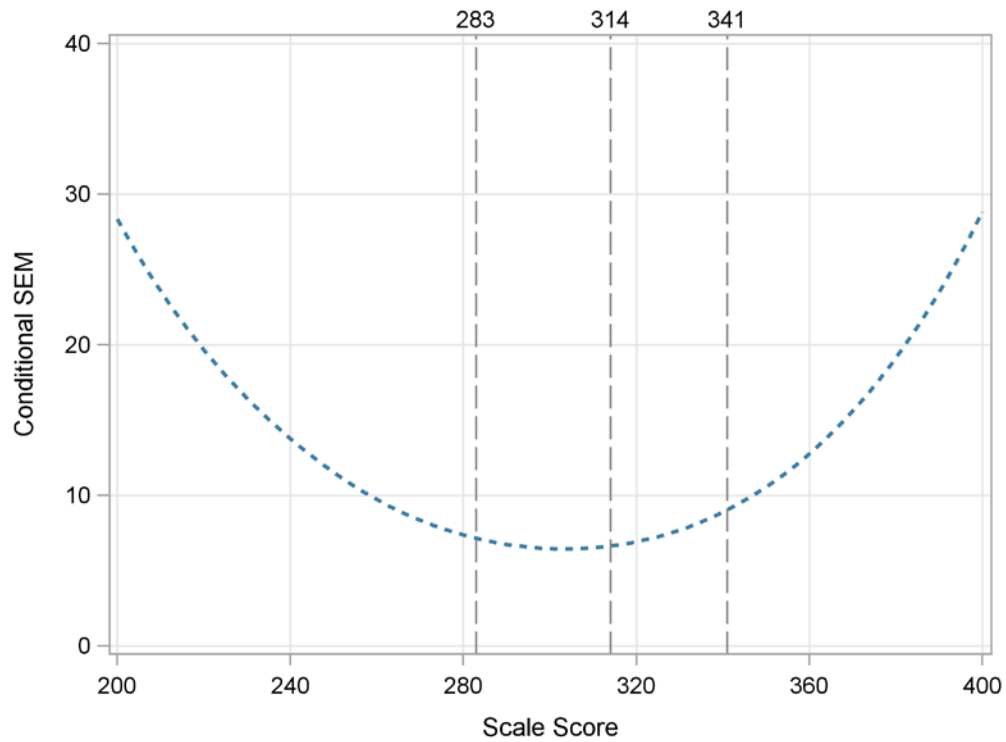


Figure 16. Mathematics Grade 4 CSEM Curve

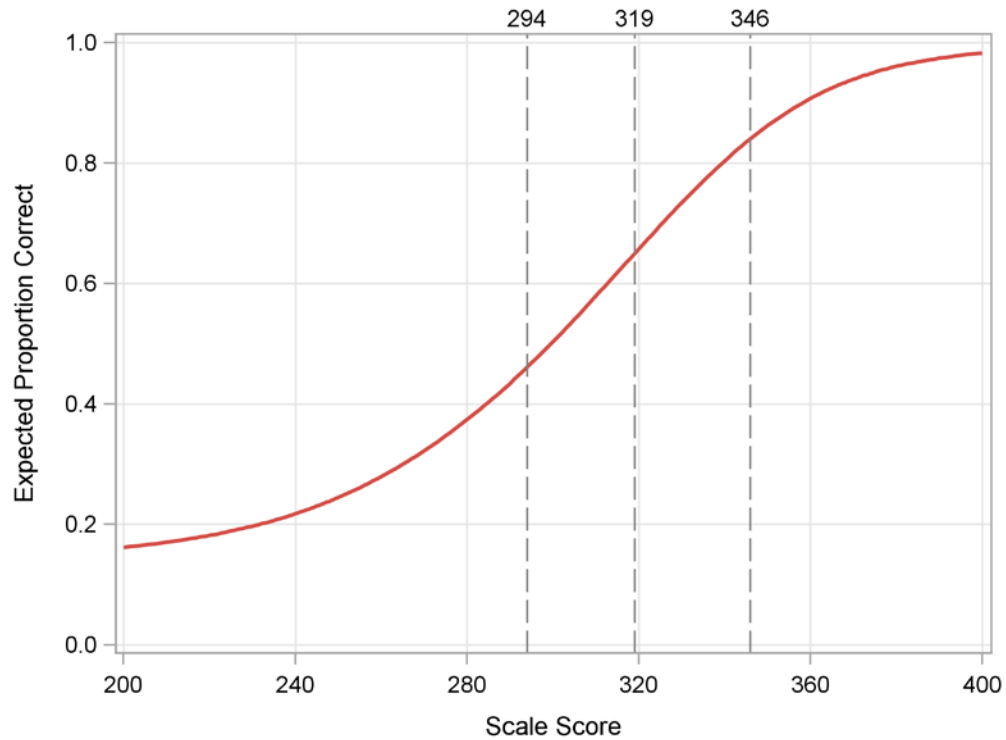


Figure 17. Mathematics Grade 5 TCC

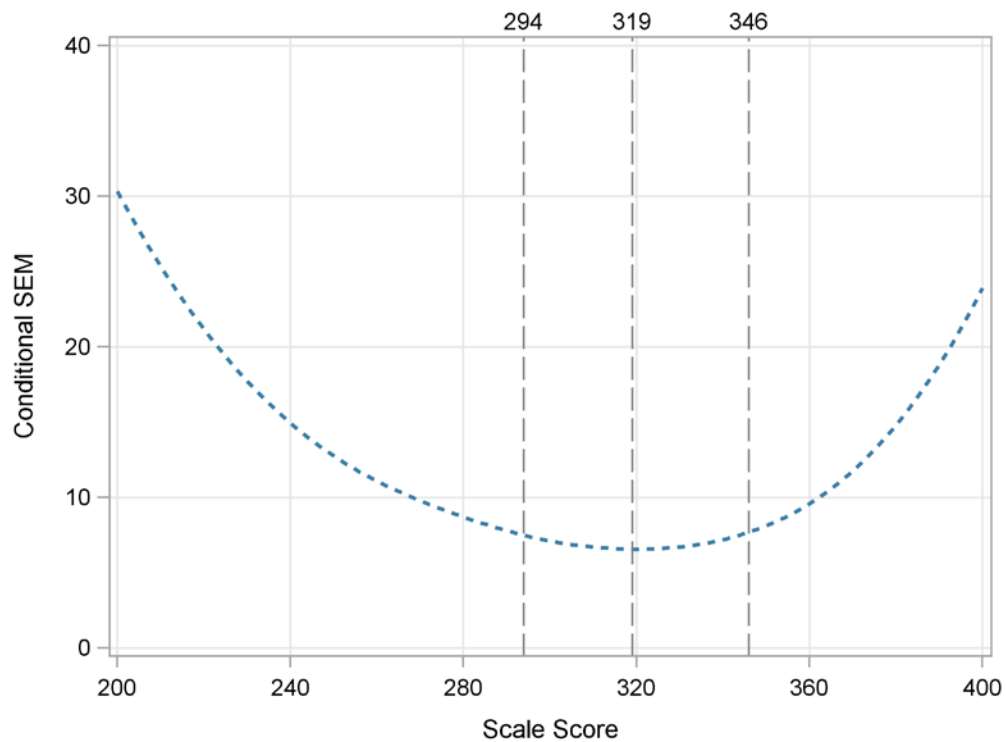


Figure 18. Mathematics Grade 5 CSEM Curve

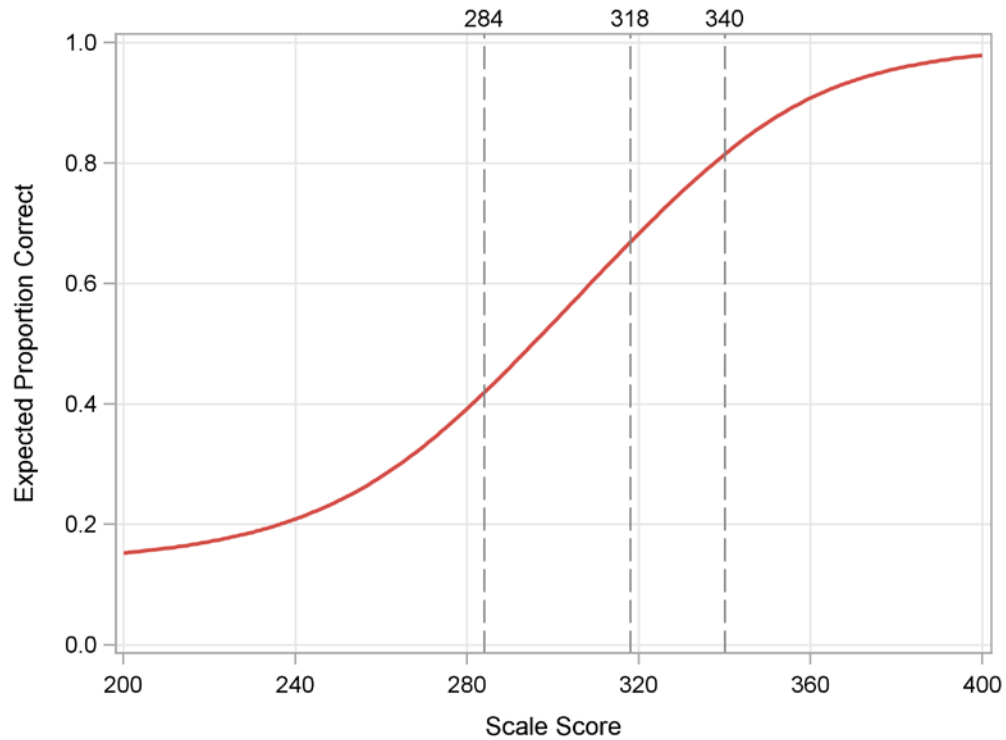


Figure 19. Mathematics Grade 6 TCC

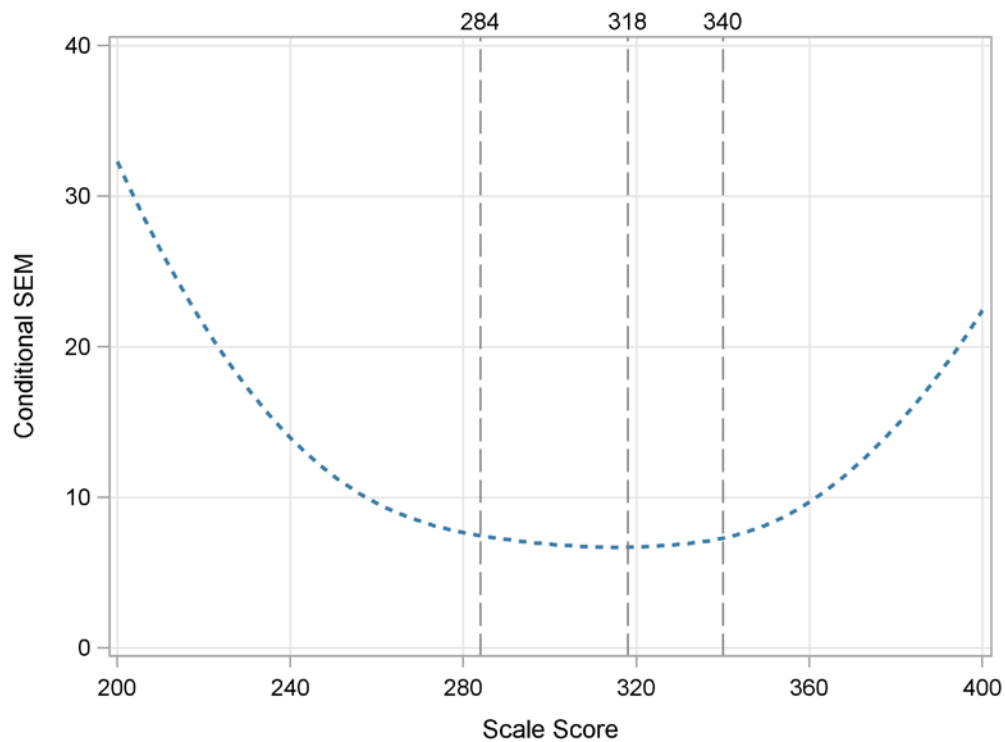


Figure 20. Mathematics Grade 6 CSEM Curve

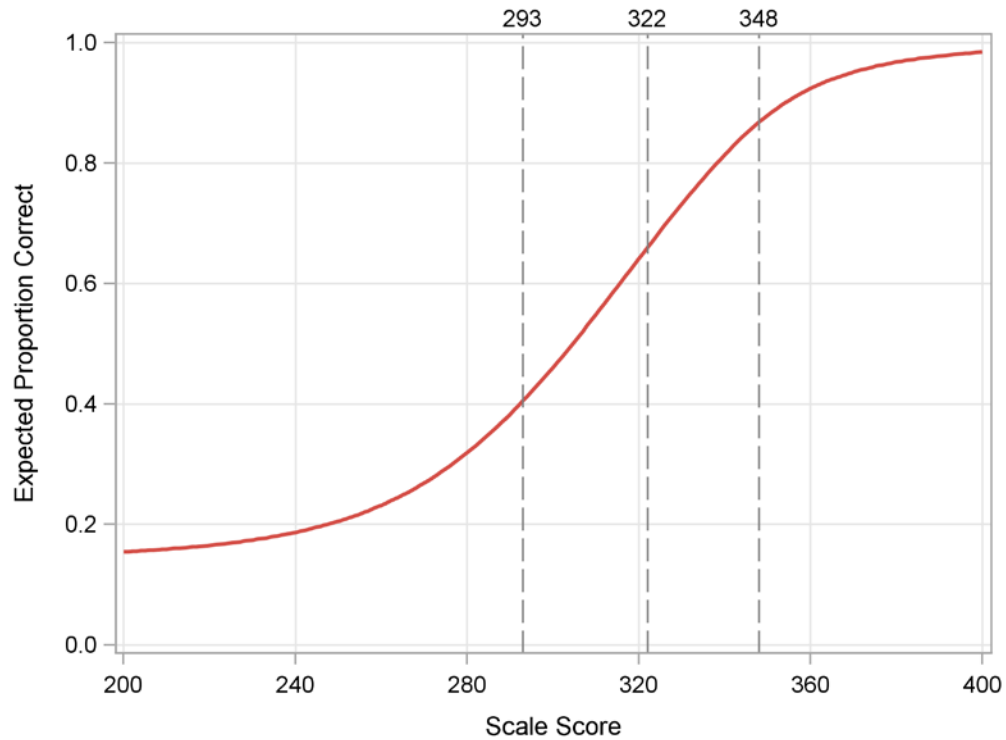


Figure 21. Mathematics Grade 7 TCC

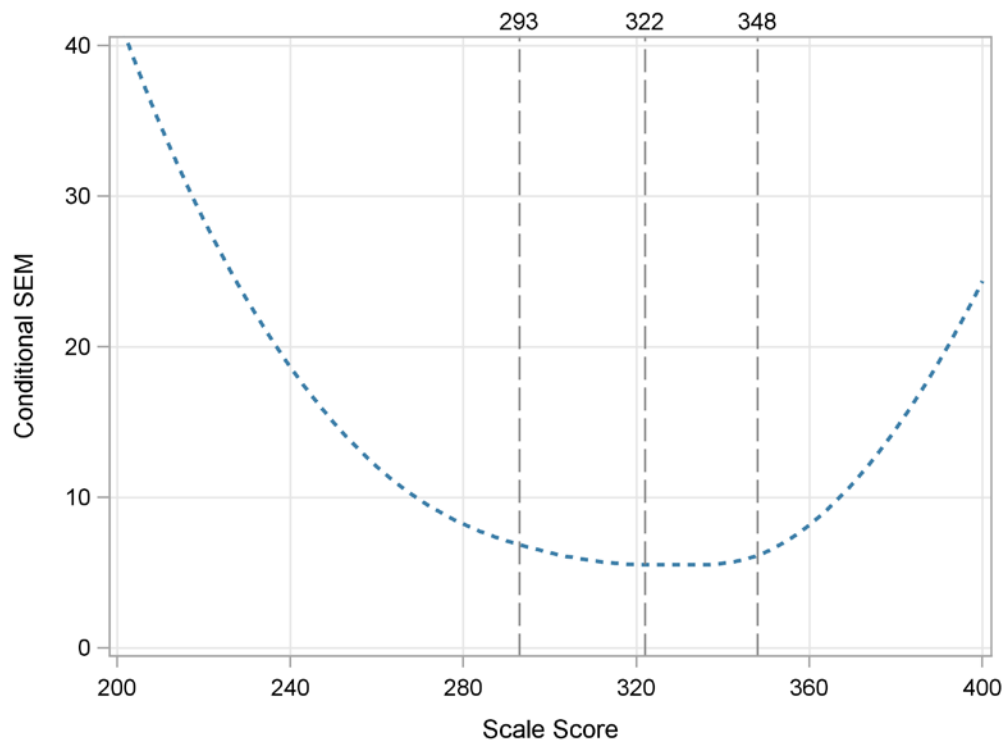


Figure 22. Mathematics Grade 7 CSEM Curve

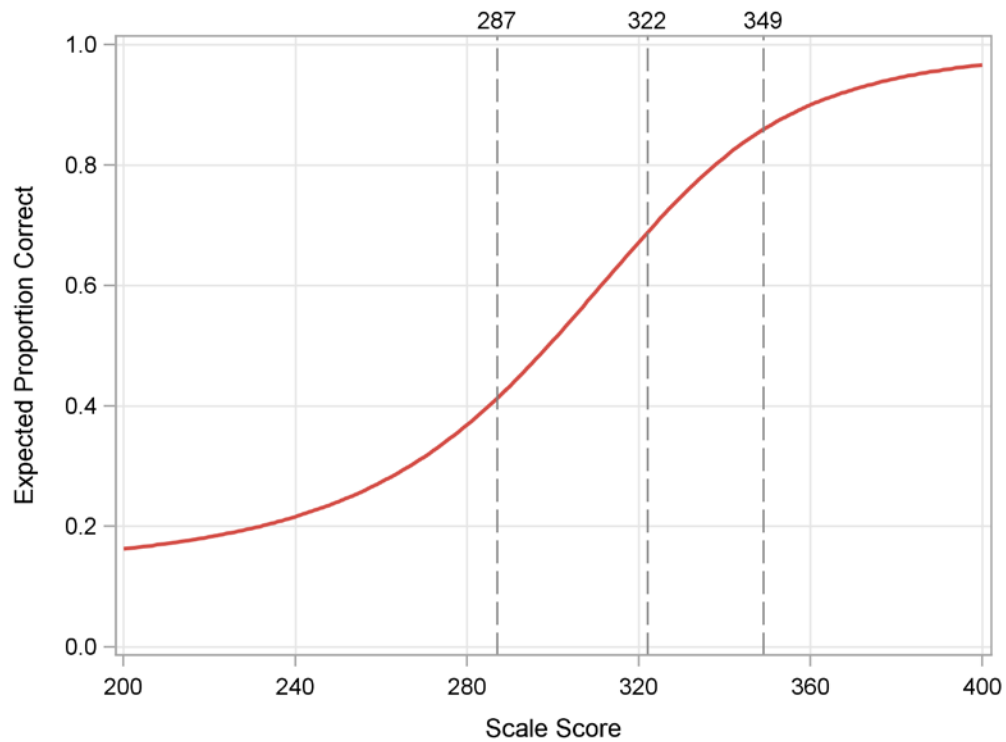


Figure 23. Mathematics Grade 8 TCC

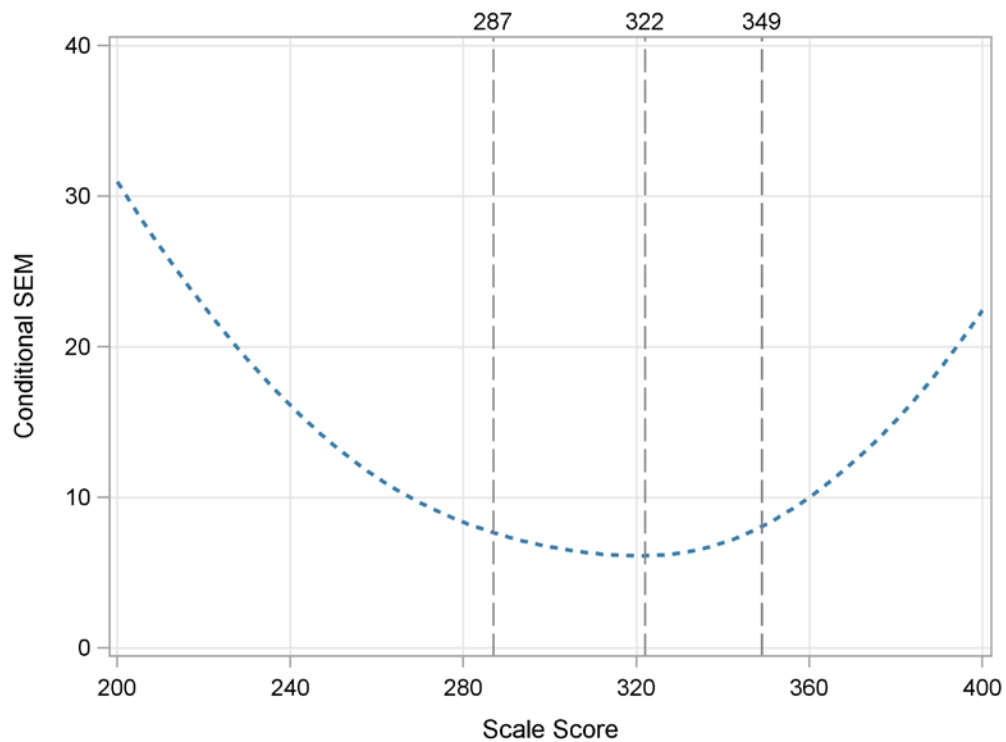


Figure 24. Mathematics Grade 8 CSEM Curve

Scoring Procedure

New York State students were scored using the number correct (NC) scoring method. This method considers how many score points a student obtained on a test in determining his or her scale score. That is, two students with the same number of score points on the test will receive the same scale score, regardless of which items they answered correctly. In this method, the number correct (or raw) score on the test is converted to a scale score by means of a conversion table. This traditional scoring method is often preferred for its conceptual simplicity and familiarity.

As described in the “Equating and Scaling” section, the final item parameters were used to calculate the raw-score-to-theta tables using a TCC method (see the details provided below). The obtained scaling transformation intercept and slope (M_1^S and M_2^S) were then applied to the theta values to produce raw score-to-scale score-conversion tables for the Grades 3–8 ELA Tests.

An inverse TCC method was employed using POLYEQUATE (Kolen and Cui, 2004). The inverse of the TCC procedure produces trait values (i.e., proficiency) based on unweighted raw scores. These estimates show negligible statistical bias (defined in statistics as the difference between an estimator’s expected value and the true value of the parameter being estimated) for tests with maximum possible raw scores of at least 30 points. All NYSTP ELA and mathematics tests have a maximum raw score higher than 30 points. In the inverse TCC method, a student’s trait (i.e., proficiency) estimate is taken to be the trait value that has an expected raw score equal to the student’s observed raw score. It was found that for tests containing only MC items, the inverse of the TCC is an excellent first-order approximation to the number of correct maximum likelihood estimates (MLE) showing negligible bias for tests of at least 30 items. For tests with a mixture of MC and CR items, the MLE and TCC estimates are even more similar (Yen, 1984).

The inverse of the TCC method relies on the following equation:

$$\sum_{i=1}^n v_i x_i = \sum_{i=1}^n v_i E(X_i | \tilde{\theta}),$$

where:

x_i is a student’s observed raw score on item i ,

v_i is a non-optimal weight specified in a scoring process ($v_i = 1$ if no weights are specified), and

$\tilde{\theta}$ is a trait estimate.

It should be noted that potential differences in test form difficulty at different proficiency levels are accounted for in the equating and in the resulting raw score-to-scale score conversion tables, so that students of the same proficiency are expected to obtain the same scale score regardless of which form they took.

Raw Score-to-Scale Score and SEM Conversion Tables

The scale score is the basic score for the NYSTP. Raw score-to-scale score (RSSS) conversion tables based on the total number correct are presented in Appendix Q, as Tables Q1–Q12.

The standard error (SE) of a scale score indicates the precision with which the proficiency is estimated, and it inversely is related to the amount of information provided by the test at each proficiency level. The SE is estimated as follows:

$$SE(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}},$$

where

$SE(\hat{\theta})$ is the standard error of the scale score (theta), and
 $I(\theta)$ is the amount of information provided by the test at a given proficiency level.

It should be noted that the information is estimated based on thetas in the scale score metric; therefore, the SE is also expressed in the scale score metric. It is also important to note that the SE value varies across proficiency levels and is the highest at the extreme ends of the scale where the amount of test information is typically the lowest. The final element of the raw-score-to-scale-score tables is the application of the performance level cut scores.

The equating procedure described above does not guarantee that the same scale score scale points selected as performance-level cut scores will be observed. It was important to appropriately reflect the performance levels set by the standard setting panel and approved by the Commissioner in the Summer of 2013. To that end, if a given scale score cut was not observed in the 2014 RSSS table, the nearest but lower scale score value was rounded up to the established scale score cut. In this way, the approved scale score cuts set in 2013 were maintained for 2014.

Table 24 and Table 25 for ELA and mathematics, respectively, show the raw- and scale-score performance level cut scores.

Table 24. ELA Performance-Level Cut Scores

Performance Level	Raw Score Cut (Scale Score Cut)					
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
NYS Level II	22 (291)	26 (287)	34 (289)	31 (283)	32 (287)	34 (284)
NYS Level III	30 (320)	36 (320)	46 (320)	48 (320)	45 (318)	48 (316)
NYS Level IV	41 (358)	43 (343)	54 (346)	54 (338)	55 (347)	57 (343)

Table 25. Mathematics Performance-Level Cut Scores

Performance Level	Raw Score Cut (Scale Score Cut)					
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
NYS Level II	29 (285)	29 (283)	30 (294)	30 (284)	29 (293)	29 (287)
NYS Level III	42 (314)	47 (314)	43 (319)	48 (318)	47 (322)	49 (322)
NYS Level IV	52 (340)	58 (341)	55 (346)	58 (340)	62 (348)	62 (349)

Section 7: Reliability and Standard Error of Measurement

This section presents specific information on various test reliability statistics and standard error of measurement (SEM), as well as the results from a study of performance level classification accuracy and consistency. The data set for these studies includes all tested New York State students who received valid scores.

Test Reliability

Test reliability is directly related to score stability and standard error and, as such, is an essential element of fairness and validity. Test reliability can be directly measured with an alpha statistic, or the alpha statistic can be used to derive the SEM. For the Grades 3–8 Common Core ELA and Mathematics Tests, we calculated two types of reliability statistics: Cronbach’s alpha (Cronbach, 1951) and Feldt-Raju coefficient (Qualls, 1995). These two measures are appropriate for assessment of a test’s internal consistency when a single test is administered to a group of examinees on one occasion. The reliability of the test is then estimated by considering how well the items that reflect the same construct yield similar results (or how consistent the results are for different items that reflect the same construct measured by the test). Both Cronbach’s alpha and Feldt-Raju coefficient measures are appropriate for tests of multiple-item formats (MC and CR items).

Test Statistics and Reliability for Total Test

Test statistics including raw-score (RS) means and raw-score standard deviations (SDs) are presented in Table 26 and Table 28 for ELA and mathematics, respectively. These statistics give the necessary context for Table 27 and Table 29, which present the case counts (n-count), number of test items (# Items), Cronbach’s alpha and associated SEM, and Feldt-Raju coefficient and associated SEM obtained for the total ELA and mathematics tests. Reliability coefficients provide measures of internal consistency that range from zero to one. High reliability indicates that scores are consistent and not unduly influenced by random error. Overall test reliability is a very good indication of each test’s internal consistency.

Grades 3–8 ELA reliability estimates (Cronbach’s alpha and Feldt-Raju) ranged from .88 to .92. Grades 3–8 mathematics reliability estimates ranged from .93 to .95 (Cronbach’s alpha and Feldt-Raju). The lowest reliabilities were observed for ELA Grade 3 and Grade 4; however, as those tests had the fewest score points, it was reasonable that its reliability would not be as high as the other grade-level tests. The highest reliability was observed for mathematics Grades 6, 7, and 8. All reliabilities were at least .88 across grades, which is a good indication that the NYSTP Grades 3–8 Common Core ELA and Mathematics Tests are acceptably reliable.

Table 26. ELA Test Form Statistics

Grade	Item-level			Student-Level			
	P-value			N-Count	Raw Score		
	Mean	Min.	Max.		Max	Mean	SD
3	0.48	0.12	0.81	198,488	49	24.95	8.81
4	0.56	0.25	0.80	203,800	55	30.32	9.63
5	0.54	0.24	0.90	199,427	66	37.85	11.89
6	0.57	0.10	0.92	200,440	66	38.83	12.61
7	0.53	0.23	0.81	197,603	65	36.52	11.91
8	0.60	0.19	0.91	203,358	66	41.35	12.14

Table 27. ELA Test Reliability and Standard Error of Measurement

Grade	N-Count	Items	Raw Score Points	Cronbach's Alpha		Feldt-Raju Coefficient	
				Est.	SEM	Est.	SEM
3	198,488	38	49	.88	3.08	.88	3.00
4	203,800	41	55	.89	3.23	.90	3.09
5	199,427	52	66	.91	3.60	.91	3.47
6	200,440	52	66	.92	3.60	.92	3.47
7	197,603	51	65	.91	3.61	.91	3.48
8	203,358	52	66	.91	3.59	.92	3.45

Table 28. Mathematics Test Form Statistics

Grade	Item-level			Student-Level			
	P-value			N-Count	Raw Score		
	Mean	Min.	Max.		Max	Mean	SD
3	0.67	0.32	0.96	200,278	60	37.39	12.78
4	0.64	0.20	0.92	204,768	66	40.35	15.98
5	0.59	0.20	0.91	198,570	66	37.68	14.36
6	0.59	0.13	0.95	199,130	72	40.82	15.75
7	0.52	0.19	0.84	193,343	72	37.54	16.36
8	0.50	0.15	0.89	159,304	72	35.82	14.62

Table 29. Mathematics Test Reliability and Standard Error of Measurement

Grade	N-Count	Items	Raw Score Points	Cronbach's Alpha		Feldt-Raju Coefficient	
				Est.	SEM	Est.	SEM
3	200,278	49	60	.93	3.47	.93	3.27
4	204,768	52	66	.95	3.74	.95	3.49
5	198,570	52	66	.94	3.63	.94	3.44
6	199,130	58	72	.94	3.84	.95	3.66
7	193,343	58	72	.94	3.95	.95	3.71
8	159,304	58	72	.93	3.95	.93	3.75

Reliability of MC Items

In addition to overall test reliability, Cronbach's alpha and Feldt-Raju coefficient were computed separately for MC and CR item sets. It is important to recognize that reliability is directly affected by test length; therefore, reliability estimates for tests by item type will always be lower than reliability estimates for the overall test form. Table 30 and Table 31 present reliabilities for the subsets of MC items.

Table 30. ELA MC Item Reliability and Standard Error of Measurement

Grade	N-Count	Items	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
3	198,488	29	.81	2.42	.81	2.42
4	203,800	31	.80	2.49	.81	2.48
5	199,427	42	.86	2.85	.86	2.84
6	200,440	42	.88	2.82	.88	2.81
7	197,603	41	.86	2.87	.86	2.86
8	203,358	42	.87	2.82	.87	2.81

Table 31. Mathematics MC Item Reliability and Standard Error of Measurement

Grade	N-Count	Items	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
3	200,278	41	.91	2.50	.91	2.47
4	204,768	42	.92	2.61	.93	2.59
5	198,570	42	.91	2.65	.91	2.63
6	199,130	48	.92	2.87	.92	2.85
7	193,343	48	.91	2.96	.91	2.95
8	159,304	48	.89	3.06	.89	3.05

Reliability of CR Items

Reliability coefficients were also computed for the subsets of CR items. The results are presented in Table 32 and Table 33.

Table 32. ELA CR Item Reliability and Standard Error of Measurement

Grade	N-Count	Items	Raw Score Points	Cronbach's Alpha		Feldt-Raju Coefficient	
				Est.	SEM	Est.	SEM
3	198,488	9	20	.82	1.75	.83	1.71
4	203,800	10	24	.86	1.83	.87	1.75
5	199,427	10	24	.85	1.96	.86	1.89
6	200,440	10	24	.86	1.98	.87	1.90
7	197,603	10	24	.87	1.93	.88	1.86
8	203,358	10	24	.86	1.95	.87	1.87

Note: Results should be interpreted with caution because the number of items is low.

Table 33. Mathematics CR Item Reliability and Standard Error of Measurement

Grade	N-Count	Items	Raw Score Points	Cronbach's Alpha		Feldt-Raju Coefficient	
				Est.	SEM	Est.	SEM
3	200,278	8	19	.82	2.16	.83	2.11
4	204,768	10	24	.88	2.36	.89	2.32
5	198,570	10	24	.85	2.27	.86	2.20
6	199,130	10	24	.86	2.31	.86	2.27
7	193,343	10	24	.89	2.26	.89	2.23
8	159,304	10	24	.86	2.25	.87	2.15

Note: Results should be interpreted with caution because the number of items is low.

Test Reliability for Reporting Categories

In this section, reliability coefficients that were estimated for the population and subgroups are presented. The reporting categories include the following: gender, ethnicity, NRC, ELL, all SWD, all SUA, students with disabilities using accommodations falling under 504 Plan (SWD/SUA), and English language learners using accommodations specific to their ELL status (ELL/SUA). Accommodations available to students under the 504 Plan include the following: Flexibility in Scheduling/Timing, Flexibility in Setting, Method of Presentation (excluding braille), Method of Response, Braille and Large-type, and others. Accommodations available to English language learners are Time Extension, Separate Location, Third Reading of Listening Selection, and Bilingual Dictionaries and Glossaries.

As shown in Table 34A–36F and Table 35 A–37F, the estimated reliabilities for subgroups were close in magnitude to the test reliability estimates of the population. Cronbach's alpha reliability coefficients were at least .80 each. Feldt-Raju reliability coefficients, which tend to be larger than the Cronbach's alpha estimates for the same group, were at least .81 each. These indicate a very good test internal consistency (reliability) for analyzed subgroups of examinees.

Table 34A. ELA Grade 3 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	198,488	.88	3.08	.88	3.00
Gender	Female	97,653	.87	3.08	.88	3.01
	Male	100,835	.88	3.07	.88	3.00
Ethnicity	Asian	17,134	.88	3.03	.89	2.95
	Black	35,369	.86	3.09	.87	3.02
	Hispanic	50,691	.85	3.08	.86	3.02
	American Indian	1,161	.86	3.11	.87	3.03
	Multiracial	3,590	.89	3.07	.90	2.98
	Pacific Islander	511	.88	3.05	.88	2.97
	White	90,032	.88	3.07	.88	2.99
NRC	New York	70,913	.88	3.07	.88	2.99
	Big 4 Cities	7,737	.86	3.05	.87	2.98
	Urban/Suburban	15,030	.85	3.08	.86	3.02
	Rural	10,327	.87	3.08	.87	3.01
	Average Needs	51,370	.87	3.08	.88	3.01
	Low Needs	24,137	.86	3.03	.87	2.97
	Charter	8,310	.85	3.11	.85	3.05
	Non-Public	10,453	.87	3.08	.88	3.01
SWD	All Codes	28,698	.85	3.03	.85	2.95
SUA	All Codes	23,906	.85	3.05	.85	2.97
ELL	ELL=Y	15,583	.81	3.07	.82	2.99
SWD/SUA	SUA=504 plan codes	13,199	.82	2.99	.83	2.92
ELL/SUA	SUA & ELL codes	5,780	.80	3.05	.81	2.98

Table 34B. ELA Grade 4 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	203,800	.89	3.23	.90	3.09
Gender	Female	100,639	.88	3.22	.89	3.09
	Male	103,161	.89	3.22	.90	3.09
Ethnicity	Asian	17,948	.88	3.12	.89	3.00
	Black	35,717	.88	3.28	.89	3.15
	Hispanic	49,427	.87	3.26	.88	3.14
	American Indian	1,125	.88	3.27	.89	3.14
	Multiracial	2,960	.90	3.20	.91	3.06
	Pacific Islander	456	.89	3.25	.90	3.11
	White	96,167	.88	3.18	.89	3.06

Table 34B. ELA Grade 4 Test Reliability by Subgroup (cont.)

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
NRC	New York	69,248	.89	3.22	.90	3.08
	Big 4 Cities	7,741	.89	3.29	.90	3.14
	Urban/Suburban	14,505	.88	3.26	.89	3.15
	Rural	10,381	.88	3.23	.89	3.12
	Average Needs	52,501	.88	3.20	.89	3.08
	Low Needs	25,985	.86	3.09	.87	3.00
	Charter	7,065	.85	3.25	.86	3.16
	Non-Public	16,166	.88	3.30	.90	3.15
SWD	All Codes	30,415	.87	3.22	.88	3.11
SUA	All Codes	26,414	.87	3.24	.88	3.13
ELL	ELL=Y	16,261	.84	3.26	.85	3.15
SWD/SUA	SUA=504 plan codes	15,723	.85	3.19	.86	3.10
ELL/SUA	SUA & ELL codes	6,423	.82	3.24	.84	3.14

Table 34C. ELA Grade 5 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	199,427	.91	3.60	.91	3.47
Gender	Female	98,168	.90	3.58	.91	3.46
	Male	101,259	.91	3.59	.92	3.47
Ethnicity	Asian	17,219	.91	3.46	.91	3.34
	Black	35,520	.90	3.66	.90	3.54
	Hispanic	47,929	.89	3.64	.90	3.53
	American Indian	1,138	.90	3.66	.91	3.52
	Multiracial	2,540	.92	3.58	.92	3.45
	Pacific Islander	457	.90	3.54	.90	3.46
	White	94,624	.91	3.55	.91	3.43
NRC	New York	67,208	.91	3.60	.91	3.47
	Big 4 Cities	7,642	.90	3.66	.90	3.54
	Urban/Suburban	14,478	.90	3.65	.90	3.54
	Rural	10,458	.90	3.63	.91	3.52
	Average Needs	54,036	.90	3.57	.91	3.46
	Low Needs	26,634	.89	3.43	.90	3.35
	Charter	8,123	.88	3.63	.89	3.55
	Non-Public	10,617	.91	3.65	.92	3.51
SWD	All Codes	31,401	.88	3.63	.89	3.53
SUA	All Codes	25,515	.88	3.63	.89	3.53
ELL	ELL=Y	13,196	.85	3.67	.86	3.55
SWD/SUA	SUA=504 plan codes	15,667	.86	3.59	.87	3.50
ELL/SUA	SUA & ELL codes	4,843	.83	3.63	.84	3.54

Table 34D. ELA Grade 6 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	200,440	.92	3.60	.92	3.47
Gender	Female	99,117	.91	3.58	.92	3.46
	Male	101,323	.92	3.59	.93	3.47
Ethnicity	Asian	16,959	.92	3.44	.92	3.31
	Black	37,103	.90	3.68	.91	3.57
	Hispanic	48,052	.90	3.67	.91	3.56
	American Indian	1,112	.91	3.65	.92	3.53
	Multiracial	2,337	.93	3.55	.93	3.41
	Pacific Islander	407	.91	3.60	.92	3.46
	White	94,470	.92	3.53	.92	3.40
NRC	New York	65,580	.92	3.63	.92	3.49
	Big 4 Cities	7,448	.91	3.67	.91	3.56
	Urban/Suburban	14,171	.91	3.65	.92	3.54
	Rural	10,497	.91	3.61	.92	3.49
	Average Needs	52,821	.91	3.55	.92	3.43
	Low Needs	26,484	.90	3.44	.91	3.32
	Charter	8,170	.89	3.64	.89	3.57
	Non-Public	15,076	.92	3.61	.92	3.48
SWD	All Codes	30,677	.89	3.63	.90	3.53
SUA	All Codes	25,894	.90	3.64	.90	3.54
ELL	ELL=Y	13,282	.86	3.68	.87	3.57
SWD/SUA	SUA=504 plan codes	16,333	.88	3.60	.89	3.51
ELL/SUA	SUA & ELL codes	5,041	.86	3.66	.87	3.56

Table 34E. ELA Grade 7 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	197,603	.91	3.61	.91	3.48
Gender	Female	97,174	.90	3.58	.91	3.46
	Male	100,429	.91	3.62	.92	3.49
Ethnicity	Asian	16,496	.91	3.48	.92	3.35
	Black	36,896	.89	3.66	.90	3.54
	Hispanic	46,537	.89	3.65	.90	3.54
	American Indian	1,096	.90	3.66	.91	3.52
	Multiracial	2,030	.92	3.59	.93	3.44
	Pacific Islander	388	.91	3.57	.92	3.43
	White	94,160	.91	3.56	.91	3.43

Table 34E. ELA Grade 7 Test Reliability by Subgroup (cont.)

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
NRC	New York	66,593	.91	3.62	.91	3.49
	Big 4 Cities	7,480	.90	3.64	.91	3.51
	Urban/Suburban	14,166	.89	3.64	.90	3.53
	Rural	10,988	.90	3.62	.91	3.49
	Average Needs	53,091	.90	3.59	.91	3.46
	Low Needs	28,019	.89	3.49	.90	3.38
	Charter	6,441	.87	3.65	.87	3.58
	Non-Public	10,603	.91	3.63	.92	3.49
SWD	All Codes	30,363	.86	3.61	.87	3.50
SUA	All Codes	21,361	.87	3.61	.88	3.49
ELL	ELL=Y	10,867	.83	3.61	.84	3.49
SWD/SUA	SUA=504 plan codes	14,243	.85	3.57	.86	3.47
ELL/SUA	SUA & ELL codes	2,816	.80	3.57	.81	3.46

Table 34F. ELA Grade 8 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	203,358	.91	3.59	.92	3.45
Gender	Female	99,644	.91	3.52	.91	3.41
	Male	103,714	.91	3.62	.92	3.49
Ethnicity	Asian	17,883	.92	3.38	.93	3.26
	Black	38,059	.90	3.69	.90	3.58
	Hispanic	47,207	.90	3.68	.90	3.57
	American Indian	1,073	.90	3.68	.90	3.56
	Multiracial	1,924	.92	3.55	.92	3.41
	Pacific Islander	432	.92	3.54	.93	3.37
	White	96,780	.91	3.49	.92	3.37
NRC	New York	68,859	.91	3.63	.92	3.50
	Big 4 Cities	7,201	.91	3.71	.91	3.58
	Urban/Suburban	13,806	.90	3.68	.91	3.57
	Rural	10,758	.91	3.62	.92	3.48
	Average Needs	53,027	.91	3.54	.91	3.42
	Low Needs	28,632	.90	3.35	.90	3.26
	Charter	5,061	.88	3.61	.88	3.54
	Non-Public	15,810	.91	3.57	.92	3.42
SWD	All Codes	29,319	.87	3.73	.88	3.62
SUA	All Codes	21,836	.89	3.72	.89	3.61
ELL	ELL=Y	10,702	.83	3.75	.85	3.63
SWD/SUA	SUA=504 plan codes	14,379	.87	3.71	.88	3.60
ELL/SUA	SUA & ELL codes	3,162	.82	3.73	.83	3.62

Table 35A. Mathematics Grade 3 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	200,278	.93	3.47	.93	3.27
Gender	Female	98,086	.92	3.46	.93	3.27
	Male	102,192	.93	3.46	.94	3.26
Ethnicity	Asian	17,655	.92	3.25	.93	3.02
	Black	35,570	.92	3.47	.93	3.33
	Hispanic	51,852	.92	3.49	.92	3.33
	American Indian	1,162	.92	3.48	.93	3.31
	Multiracial	3,594	.93	3.44	.94	3.22
	Pacific Islander	520	.93	3.41	.93	3.21
	White	89,925	.92	3.42	.93	3.23
NRC	New York	72,379	.93	3.47	.94	3.26
	Big 4 Cities	7,920	.92	3.46	.93	3.32
	Urban/Suburban	15,202	.92	3.49	.93	3.34
	Rural	10,283	.92	3.49	.93	3.32
	Average Needs	51,149	.92	3.44	.93	3.26
	Low Needs	24,253	.91	3.30	.92	3.11
	Charter	8,320	.92	3.38	.93	3.17
	Non-Public	10,561	.92	3.56	.93	3.41
SWD	All Codes	29,267	.92	3.49	.93	3.35
SUA	All Codes	33,250	.92	3.49	.93	3.35
ELL	ELL=Y	17,732	.92	3.45	.92	3.34
SWD/SUA	SUA=504 plan codes	16,588	.92	3.48	.93	3.35
ELL/SUA	SUA & ELL codes	10,133	.92	3.45	.92	3.32
ELL Test Language	English	14,377	.91	3.46	.92	3.35
	Chinese	571	.91	3.41	.92	3.23
	Haitian-Creole	51	.87	3.23	.88	3.15
	Korean	35	.87	3.25	.89	3.05
	Russian	84	.93	3.43	.94	3.22
	Spanish	2,614	.90	3.35	.90	3.28
	All Translations	3,355	.92	3.42	.93	3.29

Table 35B. Mathematics Grade 4 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	204,768	.95	3.74	.95	3.49
Gender	Female	100,866	.94	3.75	.95	3.51
	Male	103,902	.95	3.73	.95	3.47
Ethnicity	Asian	18,401	.94	3.32	.95	3.08
	Black	35,799	.94	3.79	.95	3.57
	Hispanic	50,242	.94	3.81	.95	3.59
	American Indian	1,126	.94	3.76	.95	3.51
	Multiracial	2,964	.95	3.70	.96	3.44
	Pacific Islander	466	.94	3.68	.95	3.44
	White	95,770	.94	3.69	.95	3.47
NRC	New York	70,436	.95	3.72	.96	3.44
	Big 4 Cities	7,895	.94	3.76	.95	3.55
	Urban/Suburban	14,638	.94	3.81	.95	3.59
	Rural	10,323	.94	3.82	.94	3.60
	Average Needs	52,112	.94	3.72	.95	3.50
	Low Needs	25,904	.93	3.54	.94	3.34
	Charter	7,057	.94	3.64	.94	3.40
	Non-Public	16,192	.93	3.84	.94	3.65
SWD	All Codes	30,645	.94	3.73	.94	3.54
SUA	All Codes	35,735	.94	3.78	.95	3.57
ELL	ELL=Y	18,318	.94	3.77	.94	3.57
SWD/SUA	SUA=504 plan codes	20,322	.93	3.72	.94	3.54
ELL/SUA	SUA & ELL codes	9,660	.94	3.75	.94	3.54
ELL Test Language	English	14,896	.93	3.78	.94	3.59
	Chinese	567	.93	3.60	.94	3.37
	Haitian-Creole	42	.91	3.60	.92	3.42
	Korean	32	.91	3.41	.92	3.13
	Russian	108	.94	3.79	.95	3.54
	Spanish	2,673	.91	3.65	.92	3.51
	All Translations	3,422	.94	3.73	.95	3.50

Table 35C. Mathematics Grade 5 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	198,570	.94	3.63	.94	3.44
Gender	Female	97,562	.93	3.62	.94	3.44
	Male	101,008	.94	3.63	.95	3.44
Ethnicity	Asian	17,607	.93	3.46	.94	3.19
	Black	35,284	.93	3.58	.93	3.46
	Hispanic	48,312	.93	3.60	.93	3.47
	American Indian	1,123	.94	3.60	.94	3.44
	Multiracial	2,536	.94	3.64	.95	3.44
	Pacific Islander	459	.93	3.60	.94	3.42
	White	93,249	.93	3.63	.94	3.45
NRC	New York	68,327	.94	3.61	.95	3.41
	Big 4 Cities	7,712	.92	3.55	.93	3.44
	Urban/Suburban	14,539	.92	3.59	.93	3.47
	Rural	10,314	.92	3.65	.93	3.51
	Average Needs	52,154	.93	3.64	.93	3.46
	Low Needs	26,318	.92	3.55	.93	3.35
	Charter	8,130	.93	3.59	.93	3.44
	Non-Public	10,852	.93	3.70	.94	3.55
SWD	All Codes	31,233	.91	3.53	.92	3.44
SUA	All Codes	35,184	.92	3.56	.92	3.46
ELL	ELL=Y	15,107	.92	3.55	.92	3.46
SWD/SUA	SUA=504 plan codes	21,067	.91	3.51	.91	3.43
ELL/SUA	SUA & ELL codes	7,947	.91	3.53	.91	3.44
ELL Test Language	English	11,836	.91	3.55	.92	3.46
	Chinese	528	.92	3.63	.93	3.39
	Haitian-Creole	47	.89	3.32	.90	3.23
	Korean	34	.92	3.56	.93	3.27
	Russian	81	.94	3.67	.94	3.50
	Spanish	2,581	.87	3.44	.87	3.37
	All Translations	3,271	.93	3.55	.93	3.42

Table 35D. Mathematics Grade 6 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	199,130	.94	3.84	.95	3.66
Gender	Female	98,050	.94	3.82	.94	3.65
	Male	101,080	.94	3.85	.95	3.66
Ethnicity	Asian	17,337	.94	3.55	.95	3.32
	Black	36,931	.93	3.87	.93	3.71
	Hispanic	48,743	.93	3.88	.93	3.72
	American Indian	1,076	.93	3.89	.94	3.72
	Multiracial	2,288	.94	3.85	.95	3.65
	Pacific Islander	433	.94	3.82	.95	3.61
	White	92,322	.93	3.79	.94	3.63
NRC	New York	66,787	.95	3.85	.95	3.64
	Big 4 Cities	7,543	.92	3.87	.93	3.71
	Urban/Suburban	14,154	.93	3.87	.93	3.72
	Rural	10,257	.92	3.88	.93	3.73
	Average Needs	51,112	.93	3.82	.94	3.66
	Low Needs	25,827	.93	3.64	.93	3.46
	Charter	8,150	.93	3.79	.94	3.63
	Non-Public	15,121	.93	3.91	.94	3.76
SWD	All Codes	30,321	.91	3.82	.92	3.68
SUA	All Codes	33,421	.93	3.86	.93	3.71
ELL	ELL=Y	15,253	.92	3.85	.92	3.71
SWD/SUA	SUA=504 plan codes	20,420	.91	3.81	.92	3.68
ELL/SUA	SUA & ELL codes	7,579	.92	3.82	.93	3.68
ELL Test Language	English	11,072	.91	3.86	.92	3.73
	Chinese	588	.92	3.78	.93	3.59
	Haitian-Creole	139	.85	3.65	.86	3.55
	Korean	36	.93	3.46	.94	3.22
	Russian	115	.94	3.93	.95	3.71
	Spanish	3,303	.88	3.76	.89	3.65
	All Translations	4,181	.93	3.83	.94	3.67

Table 35E. Mathematics Grade 7 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	193,343	.94	3.95	.95	3.71
Gender	Female	94,663	.94	3.95	.95	3.71
	Male	98,680	.94	3.94	.95	3.70
Ethnicity	Asian	16,670	.95	3.65	.96	3.40
	Black	36,433	.92	3.92	.93	3.74
	Hispanic	46,771	.92	3.95	.93	3.76
	American Indian	1,054	.93	3.98	.94	3.78
	Multiracial	1,932	.94	3.96	.95	3.69
	Pacific Islander	391	.95	3.83	.95	3.59
	White	90,092	.94	3.93	.94	3.71
NRC	New York	67,561	.95	3.92	.95	3.66
	Big 4 Cities	7,486	.92	3.82	.92	3.65
	Urban/Suburban	13,987	.92	3.94	.93	3.75
	Rural	10,529	.92	4.00	.93	3.81
	Average Needs	50,098	.93	3.96	.94	3.75
	Low Needs	26,445	.94	3.80	.94	3.60
	Charter	6,415	.93	3.94	.93	3.77
	Non-Public	10,616	.93	4.03	.94	3.82
SWD	All Codes	29,571	.89	3.78	.90	3.65
SUA	All Codes	24,450	.91	3.85	.92	3.69
ELL	ELL=Y	12,822	.91	3.80	.92	3.64
SWD/SUA	SUA=504 plan codes	15,559	.89	3.77	.90	3.64
ELL/SUA	SUA & ELL codes	4,417	.91	3.74	.91	3.59
ELL Test Language	English	8,752	.90	3.80	.91	3.66
	Chinese	640	.94	3.82	.94	3.59
	Haitian-Creole	138	.83	3.54	.84	3.45
	Korean	40	.93	3.68	.94	3.45
	Russian	103	.90	3.97	.91	3.75
	Spanish	3,149	.83	3.65	.84	3.55
	All Translations	4,070	.93	3.79	.93	3.59

Table 35F. Mathematics Grade 8 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	159,304	.93	3.95	.93	3.75
Gender	Female	76,750	.93	3.96	.93	3.76
	Male	82,554	.93	3.94	.94	3.74
Ethnicity	Asian	12,040	.94	3.77	.95	3.53
	Black	33,981	.91	3.89	.92	3.74
	Hispanic	42,623	.92	3.94	.92	3.76
	American Indian	892	.91	3.94	.92	3.77
	Multiracial	1,446	.93	3.95	.93	3.74
	Pacific Islander	321	.94	3.92	.94	3.69
	White	68,001	.92	3.95	.93	3.77
NRC	New York	59,190	.93	3.94	.94	3.72
	Big 4 Cities	6,812	.91	3.78	.91	3.64
	Urban/Suburban	12,110	.89	3.90	.90	3.75
	Rural	8,872	.90	3.96	.90	3.80
	Average Needs	36,583	.91	3.97	.91	3.79
	Low Needs	15,943	.92	3.87	.92	3.70
	Charter	4,861	.93	3.82	.94	3.62
	Non-Public	14,808	.93	3.97	.94	3.77
SWD	All Codes	27,318	.89	3.78	.90	3.66
SUA	All Codes	25,525	.90	3.82	.91	3.69
ELL	ELL=Y	12,437	.92	3.79	.93	3.64
SWD/SUA	SUA=504 plan codes	16,330	.88	3.77	.89	3.66
ELL/SUA	SUA & ELL codes	5,221	.91	3.72	.91	3.59
ELL Test Language	English	8,314	.91	3.77	.92	3.64
	Chinese	677	.94	3.79	.94	3.57
	Haitian-Creole	177	.81	3.62	.81	3.58
	Korean	33	.95	3.71	.95	3.43
	Russian	112	.94	3.94	.95	3.68
	Spanish	3,124	.87	3.69	.87	3.62
	All Translations	4,123	.93	3.83	.94	3.64

Standard Error of Measurement

The SEMs, as computed from Cronbach's alpha and the Feldt-Raju reliability statistics, are presented in Table 27 and Table 29 for ELA and mathematics, respectively. The SEMs ranged from 3.00 to 3.95 across subjects, grades, and the two methods of estimation, which is reasonable and small. The SEMs are directly related to reliability: the higher the reliability, the lower the standard error. As discussed, the reliability of these tests is relatively high, so it was expected that the SEMs would be very low.

The SEMs for subpopulations, as computed from Cronbach's alpha and the Feldt-Raju reliability statistics, are presented in Table 34A–36F and Table 35A–37F. The SEMs associated with all reliability estimates for all subjects, grades, methods of estimation, and subpopulations ranged from 2.92 to 4.03, which is acceptably close to those for the entire population. This narrow range indicates that across the Grades 3–8 Common Core ELA and Mathematics Tests, all students' test scores are reasonably reliable with minimal error.

Performance Level Classification Consistency and Accuracy

This subsection describes the analyses conducted to estimate performance level classification consistency and accuracy for the Grades 3–8 Common Core ELA and Mathematics Tests. In other words, this provides statistical information on the classification of students into the four performance categories. Classification consistency refers to the estimated degree of agreement between examinees' performance classification from two independent administrations of the same test (or from two parallel forms of the test). Because obtaining test scores from two independent administrations of New York State tests was not feasible due to item release after each administration, a psychometric model was used to obtain the estimated classification consistency indices using test scores from a single administration. Classification accuracy can be defined as the agreement between the actual classifications using observed cut scores and true classifications based on known true cut scores (Livingston and Lewis, 1995).

In conjunction with measures of internal consistency, classification consistency is an important type of reliability and is particularly relevant to high-stakes pass/fail tests. As a form of reliability, classification consistency represents how reliably students can be classified into performance categories.

Classification consistency is most relevant for students whose proficiency is near the pass/fail cut score. For example, consider the cut score delineating Levels II and III or simply the "Level III Cut." Students whose proficiency is far above or far below that cut score are unlikely to be misclassified because repeated administration of the test will nearly always result in the same classification. Examinees whose true scores are close to the cut score are a more serious concern. These students' true scores will likely lie within the SEM of the cut score. For this reason, the measurement error at the cut scores should be considered when evaluating the classification consistency of a test. Furthermore, the number of students near the cut scores should also be considered when evaluating classification consistency; these numbers show the number of students who are most likely to be misclassified. Scoring tables with SEMs are located in Section 6, "IRT Calibration and Scaling," and student scale score frequency distributions are located in Appendix Q. Classification consistency and accuracy were estimated using the IRT procedure

suggested by Lee, Hanson, and Brennan (2002) and Wang, Kolen, and Harris (2000). Appendix P includes a description of the calculations and procedure based on the paper by Lee et al. (2002).

Consistency

The results for classifying students into four performance levels are separated from results based solely on the Level III cut. Table 36 and Table 37 include case counts (n-count), classification consistency (Agreement), classification inconsistency (Inconsistency), and Cohen's kappa (Kappa). Consistency indicates the rate that a second administration would yield the same performance category designation (or a different designation for the inconsistency rate). The agreement index is a sum of the diagonal element in the contingency table. Kappa is similar, but corrects for chance agreement. The inconsistency index is equal to the "1 - agreement index."

Table 36 depicts the ELA and mathematics consistency study results based on the range of performance levels for all grades. For ELA, 69–74% of students were estimated to be classified consistently to one of the four performance categories with a hypothetical second administration. Kappa—which corrects for chance agreement—ranged from 0.57 to 0.62. These are between “moderate” and “substantial” agreement, as per Landis and Koch's (1977) rules of thumb for kappa. For Mathematics, 73–79% of students were estimated to be classified consistently to one of the four performance categories. Kappa, which indicates the consistency of the placement in the absence of chance, ranged from 0.64 to 0.71. These are all considered “substantial” agreement, by Landis and Koch's (1977) rules of thumb for kappa. As mentioned above and for all tests, there is an acceptable amount of measurement error that all scores contain and by random chance, students testing twice may be classified first, for example, as a Level III and second as a Level IV. This is expected to occur more often for students scoring around the selected cut score and less so for students closer to the middle of the performance level (i.e., close to the mid-point of two adjacent cut scores).

Table 36. Decision Consistency (All Cuts)

Grade	N-Count	Agreement	Inconsistency	Kappa
ELA				
3	198,488	70%	30%	0.57
4	203,800	69%	31%	0.57
5	199,427	72%	28%	0.61
6	200,440	74%	26%	0.62
7	197,603	74%	26%	0.62
8	203,358	73%	27%	0.61
Mathematics				
3	200,297	73%	27%	0.64
4	204,768	77%	23%	0.69
5	198,570	76%	24%	0.68
6	199,130	77%	23%	0.68
7	193,343	79%	21%	0.71
8	159,304	79%	21%	0.68

Table 37 depicts the ELA and mathematics consistency study results based on two performance levels (NYS Level II and NYS Level III) as defined by the Level III cut. For ELA, 88–90% of the classifications of individual students were estimated to remain stable with a second administration. Kappa coefficients for ELA classification consistency ranged from 0.72 to 0.76. These are considered “substantial” agreement, as per Landis and Koch’s (1977) rules of thumb for kappa. In addition, 90–93% of the mathematics classifications of individual students are estimated to remain stable with a second administration. Kappa coefficients for classification consistency based on the Level III cut ranged from 0.80 to 0.83. As with ELA, these statistics indicate at least “substantial” agreement (where $\text{kappa} > 0.60$) and some indicating “almost perfect” agreement (where $\text{kappa} > 0.80$), as per Landis and Koch’s (1977) rules of thumb for kappa.

Table 37. Decision Consistency (Level III Cut)

Grade	N-Count	Agreement	Inconsistency	Kappa
ELA				
3	198,488	88%	12%	0.72
4	203,800	88%	12%	0.72
5	199,427	90%	10%	0.75
6	200,440	90%	10%	0.76
7	197,603	90%	10%	0.75
8	203,358	89%	11%	0.76
Mathematics				
3	200,297	90%	10%	0.80
4	204,768	92%	8%	0.83
5	198,570	92%	8%	0.82
6	199,130	92%	8%	0.82
7	193,343	93%	7%	0.83
8	159,304	93%	7%	0.80

Accuracy

The results of classification accuracy for ELA and mathematics across all grades are presented in Table 38. Included in the table are case counts (n-count) and classification accuracy (Accuracy) for all performance levels (All Cuts) and for the Level III cut score. Note that, by definition, accuracy associated with the Level III cut is at least as great as that with the entire set of cut scores because there are only two categories for the former, as opposed to the latter, which has four.

For ELA, the estimated accuracy rates indicate that the categorization of a student’s observed performance is in agreement with the location of his or her underlying proficiency from 77% to 81% of the time across all performance levels and 91– 93% of the time in regard to the Level III cut score. For mathematics, the estimated accuracy rates indicate that the categorization of a student’s observed performance is in agreement with the location of his or her true proficiency

from 81% to 85% of the time across all performance levels and from 93% to 95% of the time in regard to the Level III cut score.

Table 38. Decision Agreement (Accuracy) Estimates

Grade	N-Count	Accuracy	
		All Cuts	Level III Cut
ELA			
3	198,488	77%	91%
4	203,800	77%	91%
5	199,427	79%	93%
6	200,440	81%	93%
7	197,603	80%	93%
8	203,358	81%	93%
Mathematics			
3	200,297	81%	93%
4	204,768	84%	94%
5	198,570	83%	94%
6	199,130	84%	94%
7	193,343	85%	94%
8	159,304	84%	95%

Section 8: Summary of Operational Test Results

This section summarizes the distribution of scale score results on the NYSTP 2014 Grades 3–8 Common Core ELA and Mathematics Tests. These include the scale score means, standard deviations, percentile ranks, and performance level distributions for each grade's population and specific subgroups. Gender, ethnic identification, NRC, ELL, SWD, and SUA variables were used to calculate the results of subgroups required for federal reporting and test equity purposes for both the ELA and mathematics tests. Additionally, the ELL/SUA subgroup is defined as English language learners who use one or more ELL-related accommodation. The SWD/SUA subgroup is defined as examinees with disabilities using one or more disability-related accommodations falling under the 504 Plan. For the mathematics analyses, the test translation language is also indicated. (Recall that the ELA tests are not translated, as they are a measure of mastery of the English language.) ELA and mathematics data include examinees with valid scores from all public, private, and charter schools. Note that complete scale score frequency distribution tables for ELA and mathematics are located in Appendix Q.

Scale Score Distribution Summary

Scale score distribution summary tables for ELA and mathematics are presented and discussed. ELA scale score distributions are described first, followed by mathematics. In the following two subsections, ELA and mathematics scale score and subscore statistics are presented for all grades, and across selected subgroups in each grade level. Use caution when interpreting the statistics for subgroups with small number counts that are included in the scale score summaries.

ELA Scale Score and Subscore Distributions

Table 39 shows some key statistics characterizing the distribution of ELA scale scores, while Table 40 summarizes the ELA subscores derived from the test in each grade. Table 41A–43F break down the scale scores by selected subgroups. Some general observations from these tables include: Females outperformed Males; Asian and White students outperformed their peers from other reported ethnic groups; students from Low Needs (as identified by NRC) districts outperformed students from other districts (New York City, Big 4 Cities, Urban/Suburban, Rural, Average Needs, and Charter); and ELL students, SWD, and/or SUA achieved below the State population (All Students) in every percentile rank. This pattern of achievement was consistent across all grades.

Table 39. ELA Scale Score Distribution Summary

Grade	N-Count	Scale Score		Percentile Ranks				
		Mean	SD	10th	25th	50th	75th	90th
3	200,504	296.80	36.73	248	273	300	321	341
4	205,221	299.11	35.13	255	280	302	325	339
5	200,909	297.36	36.41	251	276	302	322	340
6	202,022	298.30	34.51	255	277	301	322	340
7	199,911	295.00	34.76	252	273	298	320	336
8	206,140	299.25	36.03	255	277	301	324	343

Table 40. ELA Subscore Summary

Grade	Subscore	N-Count	Subscore		
			Max	Mean	SD
3	Reading	200,504	29	15.47	5.51
	Writing	200,504	20	9.37	4.17
4	Reading	205,221	31	17.73	5.64
	Writing	205,221	24	12.49	4.96
5	Reading	200,909	42	24.71	7.73
	Writing	200,909	24	13.00	5.18
6	Reading	202,022	42	25.07	8.31
	Writing	202,022	24	13.60	5.41
7	Reading	199,911	41	22.25	7.62
	Writing	199,911	24	14.07	5.39
8	Reading	206,140	42	25.62	8.00
	Writing	206,140	24	15.48	5.35

ELA Grade 3

Scale score statistics and n-counts of demographic subgroups for Grade 3 are presented in Table 41A. The population scale score mean was 296.80 with a standard deviation of 36.73. Female students tended to outperform Male students by around 9 scale score points. Asian, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students from Average Needs and Low Needs districts and Charter schools. Across reported ethnic groups, Asian students earned the highest mean score (313.66). Among NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about two-thirds of a standard deviation below the population mean. The students with disabilities (SWD), students tested under accommodations (SUA), and English language learners (ELL) subgroups scored, on average, 0.90 standard deviations below the mean scale score for the population. Students with 504 plans were the lowest-performing subgroup analyzed, scoring about 42 scale score points below the state mean. At the 50th-percentile rank, the following subgroups exceeded the population 50th-percentile rank (300): Female (304), Asian (320), Pacific Islander (304), and White (307) students and those enrolled in Average Needs (304) and Low Needs (320) districts and Charter (304) schools.

Table 41A. ELA Grade 3 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	200,504	296.80	36.73	248	273	300	321	341
Gender	Female	98,247	301.57	35.38	254	282	304	324	345
	Male	102,257	292.23	37.41	242	269	297	320	338
Ethnicity	Asian	17,198	313.66	34.85	269	293	320	338	354
	Black	35,710	286.16	35.67	242	264	291	311	331
	Hispanic	51,057	287.56	34.31	242	269	291	311	327
	American Indian	1,172	290.55	36.06	242	269	293	314	334
	Multiracial	3,663	298.43	39.81	248	273	300	327	345
	Pacific Islander	509	301.54	35.30	254	282	304	327	345
	White	91,195	302.95	36.16	254	282	307	327	345
NRC	New York	71,249	296.11	36.38	248	273	297	321	341
	Big 4 Cities	7,817	272.62	38.02	220	248	273	297	321
	Urban/Suburban	15,191	282.33	35.41	235	259	286	307	324
	Rural	10,536	286.58	36.98	235	264	291	311	331
	Average Needs	51,970	299.41	35.73	254	278	304	324	341
	Low Needs	24,445	313.24	32.68	273	293	320	334	349
	Charter	8,334	301.54	31.17	259	282	304	324	338
	Non-Public	10,745	295.94	36.75	248	278	300	321	338
SWD	All Codes	30,135	262.94	37.87	212	235	264	291	311
SUA	All Codes	24,823	267.25	38.10	212	242	269	293	314
ELL	ELL=Y	15,797	267.61	33.80	220	248	269	291	307
SWD/SUA	SUA=504 plan codes	13,791	254.86	37.25	204	228	254	282	304
ELL/SUA	SUA & ELL codes	5,893	266.16	33.62	220	242	269	291	307

ELA Grade 4

Table 41B contains Grade 4 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 299.11 with a standard deviation of 35.13. Female students tended to outperform Male students by around 8 scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students from Average Needs and Low Needs districts and Charter schools. Among reported ethnic groups, Asian students earned the highest mean score (314.94). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about two thirds of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, about one standard deviation below the mean scale score for the population. Students with 504 plans were the lowest-performing subgroup analyzed, scoring about 41 scale score points below the state mean. At the 50th-percentile rank, the following subgroups exceeded the population 50th-percentile rank (302): Female (305), Asian (320), Multiracial (305), Pacific Islander (305), and White (308) students and those enrolled in Average Needs (305) and Low Needs (315) districts and Charter (305) schools.

Table 41B. ELA Grade 4 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	205,221	299.11	35.13	255	280	302	325	339
Gender	Female	101,037	303.41	33.73	259	283	305	325	343
	Male	104,184	294.94	35.96	247	273	299	320	339
Ethnicity	Asian	17,995	314.94	33.59	273	296	320	339	356
	Black	35,967	288.54	34.50	242	266	292	311	332
	Hispanic	49,646	289.13	33.24	247	270	292	311	328
	American Indian	1,136	292.92	33.99	251	273	296	315	335
	Multiracial	3,017	302.76	37.35	255	280	305	328	347
	Pacific Islander	457	302.85	34.87	255	283	305	325	343
	White	97,003	305.14	34.01	263	287	308	328	343
NRC	New York	69,440	298.56	35.33	255	276	299	321	343
	Big 4 Cities	7,787	274.26	37.65	226	251	276	302	321
	Urban/Suburban	14,647	284.25	34.88	237	263	287	308	328
	Rural	10,546	290.80	34.79	247	270	292	315	332
	Average Needs	52,984	302.01	33.49	259	283	305	325	339
	Low Needs	26,135	312.94	30.03	276	296	315	332	347
	Charter	7,078	302.89	29.43	266	287	305	321	339
	Non-Public	16,393	298.84	36.10	251	280	302	325	339
SWD	All Codes	31,632	265.17	35.79	218	242	266	289	308
SUA	All Codes	25,259	268.30	36.04	218	247	270	292	311
ELL	ELL=Y	16,416	266.86	32.19	226	247	270	289	305
SWD/SUA	SUA=504 plan codes	14,841	258.28	35.31	210	237	259	283	302
ELL/SUA	SUA & ELL codes	5,353	262.90	31.58	218	242	266	283	302

ELA Grade 5

Scale score summary statistics by key demographic subgroups for Grade 5 students are in Table 41C. The population scale score mean was 297.36 with a standard deviation of 36.41. Female students tended to outperform Male students by around 9 scale score points. Asian, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students enrolled in New York City, Average Needs, and Low Needs districts. Across reported ethnic groups, Asian students earned the highest mean score (314.03). Among NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about three quarters of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, about one standard deviation below the mean scale score for the population. Students with 504 plans were the lowest-performing subgroup analyzed, scoring about 42 scale score points below the state mean. At the 50th-percentile rank, the following subgroups exceeded the population 50th-percentile rank (302): Female (304), Asian (317), Pacific Islander (304), and White (307) students and those from Average Needs (304) and Low Needs (317) districts.

Table 41C. ELA Grade 5 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	200,909	297.36	36.41	251	276	302	322	340
Gender	Female	98,629	302.13	34.74	258	281	304	325	346
	Male	102,280	292.76	37.37	243	270	297	320	337
Ethnicity	Asian	17,246	314.03	34.70	270	294	317	337	355
	Black	35,733	286.35	34.96	243	267	289	309	327
	Hispanic	48,140	287.96	34.24	243	267	292	312	327
	American Indian	1,149	289.90	36.87	243	270	292	314	333
	Multiracial	2,614	297.78	38.59	247	276	299	325	346
	Pacific Islander	455	300.86	33.47	258	284	304	322	340
	White	95,572	303.27	35.82	258	284	307	327	346
NRC	New York	67,460	297.63	36.06	251	276	299	322	340
	Big 4 Cities	7,698	271.04	37.16	225	247	273	297	317
	Urban/Suburban	14,600	282.73	35.59	235	261	286	307	325
	Rural	10,626	285.90	36.22	239	264	289	312	327
	Average Needs	54,620	299.56	35.02	254	278	304	322	340
	Low Needs	26,763	314.10	31.52	276	297	317	333	351
	Charter	8,137	295.31	30.62	258	276	297	317	333
	Non-Public	10,765	294.66	38.76	243	273	299	322	337
SWD	All Codes	32,515	261.83	36.68	209	239	264	286	307
SUA	All Codes	26,053	263.99	36.67	217	243	267	289	309
ELL	ELL=Y	13,304	260.16	33.57	217	239	264	284	299
SWD/SUA	SUA=504 plan codes	16,344	255.03	35.95	209	235	258	281	299
ELL/SUA	SUA & ELL codes	4,337	256.64	32.42	217	239	258	278	294

ELA Grade 6

Table 41D contains Grade 6 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 298.30 with a standard deviation of 34.51. Female students tended to outperform Male students by around 9 scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students enrolled in Average Needs and Low Needs districts. Among reported ethnic groups, Asian students earned the highest mean score (314.62). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about two thirds of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, about one standard deviation below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed, scoring about 38 scale score points below the state mean. At the 50th-percentile rank, the following subgroups exceeded the population 50th-percentile rank (301): Female (304), Asian (317), Pacific Islander (308), and White (308) students and those enrolled in Average Needs (304) and Low Needs (315) districts.

Table 41D. ELA Grade 6 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	202,022	298.30	34.51	255	277	301	322	340
Gender	Female	99,570	302.75	32.87	261	284	304	325	343
	Male	102,452	293.99	35.50	248	272	297	317	338
Ethnicity	Asian	17,002	314.62	33.47	272	295	317	338	356
	Black	37,396	286.32	32.59	245	266	288	308	325
	Hispanic	48,322	288.46	32.11	248	269	290	310	327
	American Indian	1,127	288.77	34.60	245	269	290	313	330
	Multiracial	2,377	300.45	36.14	255	279	301	325	343
	Pacific Islander	408	306.47	33.28	266	284	308	327	347
	White	95,390	305.10	33.64	261	286	308	327	343
NRC	New York	65,808	296.01	34.67	252	274	297	320	340
	Big 4 Cities	7,542	274.61	35.40	228	252	277	299	317
	Urban/Suburban	14,334	285.32	33.72	241	266	288	308	325
	Rural	10,677	292.82	33.84	248	272	297	315	333
	Average Needs	53,490	301.62	33.23	258	284	304	322	340
	Low Needs	26,508	314.44	30.06	277	297	315	333	351
	Charter	8,182	295.21	28.18	258	277	297	315	330
	Non-Public	15,273	298.08	35.01	252	279	301	320	338
SWD	All Codes	31,913	264.34	33.65	223	245	266	286	306
SUA	All Codes	25,578	268.07	34.49	223	248	269	293	310
ELL	ELL=Y	13,435	261.94	30.47	223	245	264	284	299
SWD/SUA	SUA=504 plan codes	16,447	260.23	33.79	215	241	264	284	301
ELL/SUA	SUA & ELL codes	3,987	260.00	30.42	223	241	261	283	297

ELA Grade 7

Scale score statistics and n-counts of demographic subgroups for Grade 7 are presented in Table 41E. The population scale score mean was 295.00 with a standard deviation of 34.76. Female students tended to outperform Male students by around 11 scale score points. Asian, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students from Average Needs and Low Needs districts. Across reported ethnic groups, Asian students earned the highest mean score (311.97). Among NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about two thirds of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, about one standard deviation below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed, scoring about 44 scale score points below the state mean. At the 50th-percentile rank, the following subgroups exceeded the population 50th-percentile rank (298): Female (303), Asian (318), Pacific Islander (307), and White (305) students and those enrolled in Average Needs (300) and Low Needs (314) districts.

Table 41E. ELA Grade 7 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	199,911	295.00	34.76	252	273	298	320	336
Gender	Female	97,876	300.74	32.77	258	281	303	322	340
	Male	102,035	289.50	35.72	244	268	293	314	333
Ethnicity	Asian	16,553	311.97	34.07	268	293	318	336	351
	Black	37,363	283.11	32.61	240	265	287	305	322
	Hispanic	46,881	286.28	31.97	244	268	288	307	325
	American Indian	1,115	285.66	35.31	240	265	288	310	330
	Multiracial	2,084	294.21	37.80	248	270	295	320	340
	Pacific Islander	389	304.51	34.58	262	283	307	330	347
	White	95,526	301.08	34.34	255	281	305	325	340
NRC	New York	66,827	294.50	34.06	252	273	295	318	336
	Big 4 Cities	7,587	271.12	36.05	225	248	273	295	318
	Urban/Suburban	14,432	280.39	33.58	240	262	283	303	322
	Rural	11,223	286.64	34.69	244	268	291	310	328
	Average Needs	54,002	297.01	34.01	252	278	300	320	336
	Low Needs	28,278	310.75	30.51	273	293	314	330	347
	Charter	6,459	292.92	27.78	258	276	295	312	325
	Non-Public	10,849	293.48	36.90	244	276	298	318	333
SWD	All Codes	31,848	261.39	33.26	217	240	262	283	303
SUA	All Codes	23,910	262.63	34.55	217	244	265	287	305
ELL	ELL=Y	11,017	255.71	31.14	217	240	258	276	293
SWD/SUA	SUA=504 plan codes	16,257	256.25	33.43	209	235	258	278	295
ELL/SUA	SUA & ELL codes	3,072	250.51	30.70	209	235	252	270	288

ELA Grade 8

Table 41F contains Grade 8 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 299.25 with a standard deviation of 36.03. Female students tended to outperform Male students by around 11 scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students enrolled in Average Needs and Low Needs districts and Non-Public schools. Among reported ethnic groups, Asian students earned the highest mean score (314.29). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about three quarters of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, about one standard deviation below the mean scale score for the population. English language learners tested under accommodations were the lowest performing subgroup analyzed, scoring about 49 scale score points below the state mean. At the 50th-percentile rank, the following subgroups exceeded the population 50th-percentile rank (301): Female (306), Asian (318), Multiracial (306), Pacific Islander (311), and White (308) students and those from Average Needs (306) and Low Needs (318) districts and Non-Public (306) schools.

Table 41F. ELA Grade 8 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	206,140	299.25	36.03	255	277	301	324	343
Gender	Female	100,645	304.86	34.25	261	285	306	326	347
	Male	105,495	293.90	36.86	248	272	296	318	336
Ethnicity	Asian	17,913	314.29	35.96	267	294	318	339	357
	Black	38,582	286.54	33.37	245	267	289	308	326
	Hispanic	47,589	289.37	33.21	248	269	292	311	329
	American Indian	1,096	291.30	34.90	248	272	294	316	333
	Multiracial	2,001	301.78	38.40	252	280	306	326	347
	Pacific Islander	438	306.86	38.38	258	285	311	333	352
	White	98,521	306.28	35.50	261	287	308	329	347
NRC	New York	69,167	295.55	34.72	252	275	296	318	339
	Big 4 Cities	7,330	272.89	38.36	223	252	275	299	321
	Urban/Suburban	14,111	286.66	34.63	245	267	289	308	329
	Rural	11,029	292.41	36.20	248	272	294	316	336
	Average Needs	54,214	302.68	35.25	258	284	306	326	343
	Low Needs	28,946	316.39	31.73	277	299	318	336	352
	Charter	5,067	296.61	28.04	261	280	299	316	333
	Non-Public	16,028	301.64	37.91	258	285	306	326	343
SWD	All Codes	30,866	264.02	34.68	223	245	267	287	303
SUA	All Codes	24,518	266.81	36.30	223	245	269	292	311
ELL	ELL=Y	10,832	253.25	33.18	207	233	255	275	292
SWD/SUA	SUA=504 plan codes	16,442	260.71	35.30	215	241	264	285	301
ELL/SUA	SUA & ELL codes	3,149	250.39	31.89	207	233	252	272	287

Mathematics Scale Score Distributions

Table 42 shows some key statistics characterizing the distribution of mathematics scale scores, while Table 43 summarizes the mathematics subscores derived from the test in each grade. Table 44A–44F break down the scale scores by selected subgroups. Some general observations from the mathematics data are as follows: Female and Male student performed fairly consistently; Asian students scored considerably higher than other reported ethnic groups; Low and Average Needs schools (as identified by the NRC) outperformed most other school types (New York City, Big 4 Cities, High Needs Urban/Suburban, and Rural districts), with Private and Charter Schools sometimes also outperforming other school types. Students taking the Chinese and Korean translations met or exceeded the population at most reported percentile rank, whereas the other translation subgroups (Haitian-Creole, Spanish, and Russian) were below the population scale score at each percentile rank; and ELLs, SWDs, and/or SUAs achieved below the State mean in most percentile rank ranks. This pattern of achievement was consistent across all grades.

Table 42. Mathematics Scale Score Distribution Summary

Grade	N-Count	Scale Score		Percentile Ranks				
		Mean	SD	10th	25th	50th	75th	90th
3	201,357	303.69	36.74	256	283	305	327	349
4	206,007	303.96	39.17	255	280	306	329	353
5	200,573	306.24	37.54	258	283	308	331	350
6	200,645	303.16	37.90	256	279	305	328	349
7	195,382	302.52	35.31	257	280	305	327	345
8	161,756	293.47	36.04	248	271	295	318	336

Table 43. Mathematics Subscore Summary

Grade	Subscore	N-Count	Subscore		
			Max	Mean	SD
3	Operations and Algebraic Thinking	201,357	27	15.48	6.28
	Number and Operations—Fractions	201,357	12	7.87	2.81
	Measurement and Data	201,357	12	8.44	2.68
4	Operations and Algebraic Thinking	206,007	11	5.23	3.36
	Number and Operations in Base Ten	206,007	17	10.42	4.43
	Number and Operations—Fractions	206,007	18	11.65	4.83
5	Number and Operations in Base Ten	200,573	18	11.83	3.88
	Number and Operations—Fractions	200,573	25	12.21	6.17
	Measurement and Data	200,573	10	5.44	2.69
6	Ratios and Proportional Relationships	200,645	18	9.57	4.60
	The Number System	200,645	12	7.66	2.77
	Expressions and Equations	200,645	28	15.80	6.36
7	Ratios and Proportional Relationships	195,382	20	9.89	5.19
	The Number System	195,382	14	7.54	3.60
	Expressions and Equations	195,382	22	10.76	5.11
8	Expressions and Equations	161,756	30	14.49	7.20
	Functions	161,756	19	8.92	4.04
	Geometry	161,756	12	6.71	2.57

Mathematics Grade 3

Scale score statistics and n-counts of demographic subgroups for Grade 3 are presented in Table 44A. The population scale score mean was 303.69 with a standard deviation of 36.74. Female and Male students tended to perform similarly. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students from Average Needs and Low Needs districts and Charter schools. Across reported ethnic groups, Asian students earned the highest mean score (324.58). Among NRC categories, students enrolled in Big 4 Cities districts earned the lowest mean score—by about 0.72 standard deviations below the population mean. The SWD, SUA, and ELL subgroups scored, on average, 0.72 standard deviations below the mean scale score for the population. Students with 504 plans were the lowest-performing subgroup analyzed for English forms, scoring about 33 scale score points below the state mean. At the 50th-percentile rank, the following subgroups exceeded the population 50th-percentile rank (305): Asian (325), Multiracial (307), Pacific Islander (309), and White (314) students as well as those from Average Needs (309) and Low Needs (322) districts and Charter (311) schools. In terms of the 50th-percentile ranks for students using translated forms, they ranged from 256 (Haitian-Creole, n-count = 51) to 325 (Korean, n-count = 35).

Table 44A. Mathematics Grade 3 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	201,357	303.69	36.74	256	283	305	327	349
Gender	Female	98,417	304.11	35.17	260	283	305	327	349
	Male	102,940	303.28	38.17	253	280	305	327	349
Ethnicity	Asian	17,717	324.58	34.32	283	305	325	344	370
	Black	35,789	289.19	35.47	245	266	290	314	333
	Hispanic	52,030	293.33	34.10	249	272	294	315	333
	American Indian	1,172	299.65	35.92	253	278	301	322	344
	Multiracial	3,630	305.74	39.12	256	283	307	330	354
	Pacific Islander	520	308.32	36.49	263	288	309	330	354
	White	90,499	311.23	35.11	266	292	314	333	354
NRC	New York	72,565	301.68	36.73	256	278	303	325	349
	Big 4 Cities	7,956	277.17	37.11	230	253	278	303	322
	Urban/Suburban	15,313	287.63	35.22	240	266	290	311	330
	Rural	10,407	297.84	35.52	253	278	301	320	340
	Average Needs	51,437	307.99	34.83	263	288	309	330	349
	Low Needs	24,324	321.00	32.45	283	303	322	340	361
	Charter	8,332	312.04	32.59	272	292	311	333	354
	Non-Public	10,807	299.27	36.62	253	278	303	322	340
SWD	All Codes	29,981	275.17	38.66	224	249	278	301	322
SUA	All Codes	29,116	281.01	38.11	230	256	283	307	327
ELL	ELL=Y	17,810	278.41	35.54	230	256	280	303	322
SWD/SUA	SUA=504 plan codes	15,349	270.88	38.87	216	245	272	298	318
ELL/SUA	SUA & ELL codes	8,021	279.34	35.29	236	256	280	303	322
ELL Test Language	English	14,434	278.91	35.05	236	256	280	303	322
	Chinese	571	313.04	31.65	278	296	314	333	349
	Haitian-Creole	51	252.71	35.32	208	236	256	280	292
	Korean	35	323.89	25.52	292	307	325	337	361
	Russian	84	292.49	35.85	249	266	294	315	337
	Spanish	2,635	267.63	33.04	224	245	269	290	309
	All Translations	3,376	276.29	37.46	230	253	278	301	322

Mathematics Grade 4

Table 44B contains Grade 4 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 303.96 with a standard deviation of 39.17. Female and Male students tended to perform similarly. Asian, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students enrolled in Average Needs and Low Needs districts and Charter schools. Among reported ethnic groups, Asian students earned the highest mean score (329.86). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about three quarters of a standard deviation

below the population mean. The SWD, SUA, and ELL subgroups scored, on average, 0.90 standard deviations below the mean scale score for the population. Students with 504 plans were the lowest-performing subgroup analyzed for English forms, scoring about 42 scale score points below the state mean. At the 50th percentile rank, the following groups exceeded population 50th-percentile rank (306): Asian (332), Multiracial (307), Pacific Islander (311), and White (313) students and those enrolled in Average Needs (309) and Low Needs (322) districts and Charter (314) schools. In terms of the 50th-percentile ranks for students using translated forms, they ranged from 258 (Haitian-Creole, n-count = 45) to 327 (Korean, n-count = 32).

Table 44B. Mathematics Grade 4 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	206,007	303.96	39.17	255	280	306	329	353
Gender	Female	101,235	304.26	38.03	255	281	306	329	348
	Male	104,772	303.66	40.23	251	280	306	329	353
Ethnicity	Asian	18,471	329.86	38.13	283	307	332	353	377
	Black	36,054	288.83	38.34	238	264	291	314	334
	Hispanic	50,470	292.21	37.14	243	269	294	316	337
	American Indian	1,136	297.90	39.29	251	275	297	322	348
	Multiracial	2,997	305.71	40.72	251	281	307	332	353
	Pacific Islander	468	309.18	39.00	261	285	311	336	353
	White	96,411	310.80	36.29	266	291	313	334	353
NRC	New York	70,616	303.66	41.09	251	278	304	329	353
	Big 4 Cities	7,929	275.51	39.33	224	251	275	302	324
	Urban/Suburban	14,778	285.45	38.17	238	261	287	311	332
	Rural	10,453	295.02	36.12	247	273	297	320	337
	Average Needs	52,444	307.41	35.96	261	287	309	332	348
	Low Needs	26,002	321.39	33.70	281	302	322	341	359
	Charter	7,066	314.29	35.27	271	291	314	337	359
	Non-Public	16,507	298.37	35.59	255	278	301	320	341
SWD	All Codes	31,445	271.07	38.77	216	247	271	297	320
SUA	All Codes	32,190	275.37	38.56	224	251	278	301	322
ELL	ELL=Y	18,426	275.28	37.87	224	251	275	301	322
SWD/SUA	SUA=504 plan codes	18,651	266.89	37.76	216	243	269	292	314
ELL/SUA	SUA & ELL codes	8,162	271.40	37.07	224	247	271	296	316
ELL Test Language	English	14,951	275.99	37.18	232	251	275	301	322
	Chinese	569	317.32	33.67	273	297	318	337	359
	Haitian-Creole	45	254.24	36.51	208	224	258	280	297
	Korean	32	330.47	31.04	291	309	327	348	366
	Russian	110	288.10	39.66	238	258	294	311	329
	Spanish	2,719	261.76	34.38	216	238	264	285	304
	All Translations	3,475	272.23	40.58	216	247	271	299	324

Mathematics Grade 5

Grade 5 demographic subgroup n-counts and scale score statistics are presented in Table 44C. The population scale score mean was 306.24 with a standard deviation of 37.54. Female and male students tended to perform similarly. Asian, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students from Average Needs and Low Needs districts and Charter schools. Across reported ethnic groups, Asian students earned the highest mean score (332.39). Among NRC categories, students enrolled in Big 4 Cities districts earned the lowest mean score—by about four fifths of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, about 0.85 standard deviations below the mean scale score for the population. Students with 504 plans were the lowest-performing subgroup analyzed for English forms, scoring about 37 scale score points below the state mean. At the 50th-percentile rank, the following subgroups exceeded the population 50th-percentile rank (308): Female (310), Asian (333), Pacific Islander (314), and White (315) students and those from Average Needs (311), Low Needs (327) districts. In terms of the 50th-percentile ranks for students using translated forms, they ranged from 262 (Haitian-Creole, n-count = 49) to 323 (Korean, n-count = 35).

Table 44C. Mathematics Grade 5 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	200,573	306.24	37.54	258	283	308	331	350
Gender	Female	98,329	307.26	35.95	262	286	310	331	350
	Male	102,244	305.26	38.99	254	281	308	331	353
Ethnicity	Asian	17,680	332.39	36.01	288	311	333	357	377
	Black	35,684	291.16	35.79	244	269	294	315	335
	Hispanic	48,900	295.82	35.01	249	275	297	319	338
	American Indian	1,134	300.85	38.30	249	278	302	325	346
	Multiracial	2,578	305.43	39.25	254	281	308	333	353
	Pacific Islander	466	310.23	37.94	266	291	314	335	353
	White	94,131	312.52	35.49	266	294	315	335	353
NRC	New York	68,577	307.01	38.48	258	283	308	333	353
	Big 4 Cities	7,754	277.32	38.06	230	254	278	304	327
	Urban/Suburban	14,641	287.69	35.54	244	266	291	311	331
	Rural	10,455	296.79	34.53	254	278	299	319	338
	Average Needs	53,329	308.81	34.68	266	288	311	331	350
	Low Needs	26,388	324.70	31.98	286	306	327	346	365
	Charter	8,137	306.95	32.72	266	286	308	329	347
	Non-Public	11,059	298.36	38.39	249	275	302	325	342
SWD	All Codes	32,072	274.02	36.96	222	249	275	299	319
SUA	All Codes	29,176	277.58	36.98	230	254	281	304	323
ELL	ELL=Y	15,257	276.11	36.84	230	254	278	299	321
SWD/SUA	SUA=504 plan codes	17,483	269.54	35.98	222	244	272	295	315
ELL/SUA	SUA & ELL codes	6,204	272.60	35.48	222	249	275	295	315

Table 44C. Mathematics Grade 5 Scale Score Distribution by Subgroup (cont.)

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
ELL Test Language	English	11,931	276.40	35.94	230	254	278	299	319
	Chinese	530	322.31	32.27	282	302	321	346	360
	Haitian-Creole	49	252.86	39.35	198	222	262	275	302
	Korean	35	326.31	32.06	286	302	323	357	370
	Russian	81	294.02	40.13	244	266	299	325	340
	Spanish	2,631	264.71	33.14	222	244	269	288	304
	All Translations	3,326	275.07	39.88	222	249	275	299	325

Mathematics Grade 6

Table 44D contains Grade 6 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 303.16 with a standard deviation of 37.90. Female students tended to outperform Male students by around 3 scale score points. Asian, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students enrolled in Average Needs and Low Needs districts and Charter schools. Among reported ethnic groups, Asian students earned the highest mean score (330.39). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about 0.75 standard deviations below the population mean. The SWD, SUA, and ELL subgroups scored, on average, 0.85 standard deviations below the mean scale score for the population. Students with 504 plans were the lowest-performing subgroup analyzed for English forms, scoring about 36 scale score points below the state mean. At the 50th-percentile rank, the following subgroups exceeded the population 50th-percentile rank (305): Female (306), Asian (332), Pacific Islander (310), and White (312) students and those enrolled in Average Needs (308), and Low Needs (326) and Charter (308) schools. In terms of the 50th-percentile ranks for students using translated forms, they ranged from 262 (Haitian-Creole, n-count = 140) to 340 (Korean, n-count = 36).

Table 44D. Mathematics Grade 6 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	200,645	303.16	37.90	256	279	305	328	349
Gender	Female	98,538	304.72	36.48	260	282	306	330	349
	Male	102,107	301.66	39.16	253	277	303	328	349
Ethnicity	Asian	17,426	330.39	37.59	284	308	332	355	373
	Black	37,231	287.19	35.72	240	265	288	310	332
	Hispanic	49,019	291.10	35.22	245	270	293	314	334
	American Indian	1,095	292.94	36.88	245	270	295	316	340
	Multiracial	2,319	302.93	38.75	253	279	305	328	349
	Pacific Islander	433	309.01	39.06	260	286	310	334	355
	White	93,122	310.90	35.04	268	290	312	334	352

Table 44D. Mathematics Grade 6 Scale Score Distribution by Subgroup (cont.)

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
NRC	New York	67,002	300.38	39.98	249	275	301	328	352
	Big 4 Cities	7,628	275.74	37.26	226	253	277	301	322
	Urban/Suburban	14,305	286.38	35.27	240	265	288	310	330
	Rural	10,431	295.45	34.03	253	275	297	318	336
	Average Needs	51,565	306.97	34.40	263	288	308	330	349
	Low Needs	25,929	324.24	32.56	284	305	326	346	363
	Charter	8,161	307.58	32.71	265	288	308	330	346
	Non-Public	15,434	299.21	35.57	253	279	301	322	341
SWD	All Codes	31,225	268.89	36.32	218	245	270	293	314
SUA	All Codes	28,565	275.29	37.28	226	253	277	301	322
ELL	ELL=Y	15,360	271.10	36.68	226	249	270	293	316
SWD/SUA	SUA=504 plan codes	17,488	267.27	35.96	218	245	268	292	312
ELL/SUA	SUA & ELL codes	5,755	270.50	37.31	226	249	270	293	318
ELL Test Language	English	11,130	271.12	35.73	226	249	273	293	316
	Chinese	600	319.88	31.57	277	301	320	341	358
	Haitian-Creole	140	255.03	32.16	210	237	262	275	295
	Korean	36	332.33	30.52	284	314	340	351	367
	Russian	119	287.43	38.85	240	263	286	312	341
	Spanish	3,335	261.70	32.77	218	240	263	284	303
	All Translations	4,230	271.06	39.08	218	245	270	297	321

Mathematics Grade 7

The n-counts and scale score statistics for key demographic subgroups of Grade 7 students are presented in Table 44E. The population scale score mean was 302.52 with a standard deviation of 35.31. Female students tended to outperform Male students by around 3 scale score points. Asian, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students from Average Needs and Low Needs districts and Charter schools. Across reported ethnic groups, Asian students earned the highest mean score (326.67). Among NRC categories, students enrolled in Big 4 Cities districts earned the lowest mean score—by about three quarters of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, 0.90 standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 36 scale score points below the state mean. At the 50th-percentile rank, the following subgroups exceeded the population 50th-percentile rank (305): Female (306), Asian (330), Pacific Islander (317), and White (314) students and those from Average Needs (309) and Low Needs (324) districts and Charter (306) schools. In terms of the 50th-percentile ranks for students using translated forms, they ranged from 257 (Haitian-Creole, n-count = 141) to 324 (Chinese, n-count = 647).

Table 44E. Mathematics Grade 7 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	195,382	302.52	35.31	257	280	305	327	345
Gender	Female	95,301	304.02	33.94	261	282	306	327	345
	Male	100,081	301.08	36.51	257	278	303	326	345
Ethnicity	Asian	16,746	326.67	34.26	282	308	330	349	367
	Black	36,907	286.26	33.22	244	265	289	309	327
	Hispanic	47,146	291.07	32.61	251	272	293	314	330
	American Indian	1,079	293.80	34.86	251	275	296	317	335
	Multiracial	1,976	302.75	36.21	257	280	303	327	348
	Pacific Islander	392	314.61	35.16	272	293	317	339	358
	White	91,136	310.63	32.59	269	293	314	332	349
NRC	New York	67,826	300.76	36.69	257	278	301	326	348
	Big 4 Cities	7,575	274.45	35.02	228	251	275	298	318
	Urban/Suburban	14,240	285.55	32.84	244	265	287	308	326
	Rural	10,784	295.69	31.12	257	278	298	317	332
	Average Needs	50,792	306.28	32.42	265	289	309	329	343
	Low Needs	26,588	320.91	30.36	282	305	324	341	355
	Charter	6,427	304.29	29.81	269	287	306	324	339
	Non-Public	10,910	298.77	33.95	257	280	301	322	339
SWD	All Codes	30,757	270.00	32.99	228	251	272	293	311
SUA	All Codes	24,654	274.16	34.42	228	251	275	298	317
ELL	ELL=Y	12,962	270.74	34.85	228	251	272	293	314
SWD/SUA	SUA=504 plan codes	15,926	268.18	33.02	220	244	269	291	309
ELL/SUA	SUA & ELL codes	4,271	266.26	34.84	220	244	265	287	309
ELL Test Language	English	8,821	270.10	33.82	228	251	272	291	312
	Chinese	647	319.48	29.66	280	303	324	339	352
	Haitian-Creole	141	255.50	31.52	212	235	257	275	298
	Korean	40	323.85	26.17	292	313	323	343	354
	Russian	105	285.20	28.05	244	265	285	306	322
	Spanish	3,208	262.21	30.04	220	244	265	282	298
	All Translations	4,141	272.11	36.91	220	251	272	295	323

Mathematics Grade 8

It should be noted that students in the 2013 population did not have available to them the double testing waiver and hence included more students who were accelerated and taking the Integrated Algebra Regents Examination. So the 2014 population—summarized in Table 44F—was smaller (n-count = 161,756) than the 2013 population (n-count = 214,932) and had slightly lower (mean = 293.47, as opposed to 299.43) and somewhat more variable (SD = 36.04, as opposed to 35.23) scale scores.

That table also contains Grade 8 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 293.47 with a standard deviation of 36.04. Female students tended to outperform Male students by around 4 scale score points. Asian, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students enrolled in New York City, Average Needs, and Low Needs districts and Charter and Non-Public schools. Among reported ethnic groups, Asian students earned the highest mean score (318.96). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about 0.75 standard deviations below the population mean. The SWD, SUA, and ELL subgroups scored, on average, scored 0.76 standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest performing subgroup analyzed for English forms, scoring about 35 scale score points below the state mean. At the 50th-percentile rank, the following subgroups exceeded the population 50th-percentile rank (295): Female (297), Asian (322), Pacific Islander (301), and White (302) students and those enrolled in Average Needs (299) and Low Needs (313) districts and Charter (307) and Non-Public (299) schools. In terms of the 50th-percentile ranks for students using translated forms, they ranged from 261 (Haitian-Creole, n-count = 183) to 326 (Korean, n-count = 34).

Table 44F. Mathematics Grade 8 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	161,756	293.47	36.04	248	271	295	318	336
Gender	Female	77,575	295.39	34.88	253	274	297	318	336
	Male	84,181	291.70	36.99	243	271	295	316	334
Ethnicity	Asian	12,099	318.96	38.64	271	297	322	344	366
	Black	34,585	280.91	34.97	236	261	283	304	323
	Hispanic	43,081	286.17	34.46	243	265	288	309	326
	American Indian	909	284.55	34.07	243	268	287	307	326
	Multiracial	1,494	292.63	36.81	243	271	295	318	336
	Pacific Islander	325	298.72	36.54	253	274	301	326	341
	White	69,263	299.93	33.23	257	283	302	322	336
NRC	New York	59,604	293.64	37.52	248	271	293	318	339
	Big 4 Cities	6,967	265.93	36.75	220	243	268	290	313
	Urban/Suburban	12,408	277.27	32.48	236	257	281	299	316
	Rural	9,106	286.84	31.06	248	271	290	307	323
	Average Needs	37,396	295.60	31.63	257	279	299	316	330
	Low Needs	16,124	309.58	31.20	271	293	313	328	344
	Charter	4,870	305.89	33.83	261	283	307	328	346
	Non-Public	15,110	296.65	37.82	248	276	299	322	341
SWD	All Codes	28,623	265.23	34.86	220	243	268	288	307
SUA	All Codes	20,867	268.55	36.19	220	248	271	293	313
ELL	ELL=Y	12,646	267.86	38.59	220	243	268	292	316
SWD/SUA	SUA=504 plan codes	13,887	263.52	35.01	212	243	265	288	306
ELL/SUA	SUA & ELL codes	3,215	258.30	37.16	212	236	257	281	304

Table 44F. Mathematics Grade 8 Scale Score Distribution by Subgroup (cont.)

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
ELL Test Language	English	8,445	265.77	37.37	220	243	265	290	313
	Chinese	694	320.53	36.74	276	299	323	341	366
	Haitian-Creole	183	256.95	29.75	212	236	261	281	292
	Korean	34	321.88	37.34	268	297	326	344	357
	Russian	114	288.19	39.14	236	265	293	311	339
	Spanish	3,176	261.24	32.95	212	243	265	283	301
	All Translations	4,201	272.07	40.60	220	248	271	297	325

Performance Level Distribution Summary

Students are classified as NYS Level I, NYS Level II, NYS Level III, and NYS Level IV. The cut scores were established in 2013 during the standard-setting. Table 24 and Table 25 show the ELA and mathematics cut scores, respectively, used for classification of students into the four performance-level categories in 2014. Please note that it is inappropriate to compare scale scores across grades as they neither measure the same content, nor are they on the same scale. During the standard-setting process, while cut scores were set separately for different grades within a subject, additional care was taken to vertically articulate performance levels; see Section 8 and Appendix P in the 2013 technical report (NYSED, 2014) for details. While vertical articulation helps to build consistent meaning to the performance levels, the very nature of grade-specific content, differing performance expectations, and panel set cut scores result in cut score differences across grades.

ELA Test Performance Level Distributions

Table 45 shows the performance level distribution for all examinees from public, charter, and private schools with valid ELA scores. Performance level data for selected subgroups of students were also examined. In general, these distributions reflect the same achievement trends in the scale score summary discussion. More Female students were classified in Level III and above categories than Male students. Similarly, more Asian and White students were classified in Level III and above categories than their peers from other reported ethnic groups. Consistent with the pattern shown in scale score distribution across the subgroups, students from Low and Average Needs districts outperformed students from High Needs districts (New York City, Big 4 Cities, Urban/Suburban, and Rural). The Level III and above rates for students in the ELL, SWD, and SUA subgroups were low compared to the total population of examinees.

Table 45. ELA Test Performance Level Distributions

Grade	N-Count	Level I	Level II	Level III	Level IV	Level III & IV
3	200,504	37.07	31.34	27.96	3.63	31.59
4	205,221	30.68	36.83	22.69	9.79	32.49
5	200,909	35.92	35.05	19.78	9.25	29.03
6	202,022	27.17	44.83	15.12	12.89	28.01
7	199,911	35.33	36.26	22.63	5.78	28.41
8	206,140	27.46	37.93	24.18	10.42	34.60

ELA Grade 3

Performance level distributions and n-counts of demographic subgroups for ELA Grade 3 are presented in Table 46A. Statewide, a combined 32% of students achieved Level III and Level IV. About 36% of Female students were at Level III or above, as compared to 28% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 51% each of Asian students and students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 12–21% of students in those same performance categories. Only 7% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (32%): Female (36%), Asian (51%), Multiracial (36%), Pacific Islander (37%), and White (39%) students and those from Average Needs (34%) and Low Needs (51%) districts and Charter (34%) schools.

Table 46A. ELA Grade 3 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	200,504	37.07	31.34	27.96	3.63	31.59
Gender	Female	98,247	32.32	31.86	31.18	4.65	35.82
	Male	102,257	41.63	30.83	24.87	2.66	27.53
Ethnicity	Asian	17,198	20.87	28.06	41.98	9.09	51.07
	Black	35,710	49.05	30.67	18.84	1.43	20.28
	Hispanic	51,057	47.41	32.57	18.66	1.36	20.02
	American Indian	1,172	43.43	32.25	22.27	2.05	24.32
	Multiracial	3,663	35.93	28.47	30.41	5.19	35.60
	Pacific Islander	509	33.40	29.47	33.40	3.73	37.13
	White	91,195	29.63	31.64	34.04	4.70	38.73
NRC	New York	71,249	38.95	31.07	26.11	3.87	29.98
	Big 4 Cities	7,817	63.87	23.74	11.47	0.91	12.38
	Urban/Suburban	15,191	53.58	29.83	15.52	1.07	16.59
	Rural	10,536	47.39	31.51	19.40	1.70	21.10
	Average Needs	51,970	33.17	32.67	30.57	3.59	34.16
	Low Needs	24,445	19.17	30.30	43.42	7.11	50.53
	Charter	8,334	30.35	35.49	31.44	2.72	34.16
	Non-Public	10,745	36.39	33.20	27.78	2.63	30.41
SWD	All Codes	30,135	73.66	18.74	7.14	0.46	7.60
SUA	All Codes	24,823	68.72	22.03	8.76	0.49	9.25
ELL	ELL=Y	13,791	80.33	14.89	4.60	0.19	4.79
SWD/SUA	SUA=504 plan codes	5,893	73.27	20.86	5.68	0.19	5.87
ELL/SUA	SUA & ELL codes	15,797	71.77	21.83	6.21	0.19	6.40

ELA Grade 4

Performance level distributions and n-counts of demographic subgroups for ELA Grade 4 are presented in Table 46B. Statewide, a combined 32% of students achieved Level III and Level IV. About 37% of Female students were at Level III or above, as compared to 29% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 52% of Asian students and 48% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 13–23% of students in those same performance categories. Only 6% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (32%): Female (37%), Asian (52%), Multiracial (38%), Pacific Islander (34%), and White (39%) students and those enrolled in Average Needs (35%) and Low Needs (48%) districts.

Table 46B. ELA Grade 4 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	205,221	30.68	36.83	22.69	9.79	32.49
Gender	Female	101,037	26.48	36.94	24.76	11.83	36.59
	Male	104,184	34.77	36.72	20.69	7.82	28.51
Ethnicity	Asian	17,995	16.07	32.13	30.81	20.99	51.80
	Black	35,967	42.37	36.76	16.16	4.72	20.87
	Hispanic	49,646	41.49	38.22	16.07	4.22	20.29
	American Indian	1,136	37.59	38.56	16.99	6.87	23.86
	Multiracial	3,017	28.27	33.97	23.40	14.35	37.75
	Pacific Islander	457	27.79	37.86	21.88	12.47	34.35
	White	97,003	23.54	37.08	27.05	12.34	39.38
NRC	New York	69,440	32.62	36.15	20.76	10.47	31.23
	Big 4 Cities	7,787	58.87	28.16	9.99	2.98	12.97
	Urban/Suburban	14,647	47.53	35.01	13.89	3.57	17.46
	Rural	10,546	38.88	38.22	17.52	5.38	22.90
	Average Needs	52,984	26.33	38.63	25.10	9.94	35.04
	Low Needs	26,135	14.90	36.82	32.89	15.39	48.28
	Charter	7,078	24.84	41.90	24.84	8.42	33.26
	Non-Public	16,393	30.51	36.45	23.17	9.88	33.04
SWD	All Codes	31,632	69.73	23.46	5.48	1.34	6.81
SUA	All Codes	25,259	65.70	26.13	6.87	1.30	8.17
ELL	ELL=Y	14,841	76.32	19.49	3.62	0.57	4.19
SWD/SUA	SUA=504 plan codes	5,353	75.06	21.74	2.88	0.32	3.19
ELL/SUA	SUA & ELL codes	16,416	70.08	24.98	4.33	0.60	4.93

ELA Grade 5

Performance level distributions and n-counts of demographic subgroups for ELA Grade 5 are presented in Table 46C. Statewide, a combined 29% of students achieved Level III and Level IV. About 33% of Female students were at Level III or above, as compared to 25% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 48% each of Asian students and students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 10–18% of students in those same performance categories. Only 4% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (29%): Female (33%), Asian (48%), Multiracial (31%), Pacific Islander (31%), and White (36%) students and those from Average Needs (31%) and Low Needs (48%) districts.

Table 46C. ELA Grade 5 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	200,909	35.92	35.05	19.78	9.25	29.03
Gender	Female	98,629	31.02	35.93	21.82	11.23	33.06
	Male	102,280	40.65	34.21	17.80	7.34	25.14
Ethnicity	Asian	17,246	18.97	32.74	28.13	20.15	48.28
	Black	35,733	48.64	34.39	12.90	4.07	16.97
	Hispanic	48,140	46.46	35.78	13.77	3.99	17.76
	American Indian	1,149	43.17	35.42	15.75	5.66	21.41
	Multiracial	2,614	37.30	31.33	19.20	12.17	31.37
	Pacific Islander	455	28.79	40.44	22.86	7.91	30.77
	White	95,572	28.82	35.43	23.91	11.84	35.75
NRC	New York	67,460	36.46	35.07	18.73	9.74	28.47
	Big 4 Cities	7,698	66.76	23.71	7.56	1.97	9.53
	Urban/Suburban	14,600	52.39	32.66	11.73	3.22	14.95
	Rural	10,626	48.76	33.23	13.95	4.07	18.01
	Average Needs	54,620	32.61	36.89	21.39	9.11	30.50
	Low Needs	26,763	17.28	35.11	30.13	17.48	47.61
	Charter	8,137	37.96	39.92	17.06	5.06	22.12
	Non-Public	10,765	36.91	35.02	19.94	8.13	28.07
SWD	All Codes	32,515	75.97	19.06	4.07	0.90	4.97
SUA	All Codes	26,053	73.34	21.17	4.60	0.89	5.48
ELL	ELL=Y	16,344	81.91	15.31	2.44	0.34	2.78
SWD/SUA	SUA=504 plan codes	4,337	84.55	13.67	1.50	0.28	1.78
ELL/SUA	SUA & ELL codes	13,304	79.81	17.78	2.00	0.41	2.41

ELA Grade 6

Performance level distributions and n-counts of demographic subgroups for ELA Grade 6 are presented in Table 46D. Statewide, a combined 28% of students achieved Level III and Level IV. About 32% of Female students were at Level III or above, as compared to 24% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 48% of Asian students and 46% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 9–22% of students in those same performance categories. Only 4% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (28%): Female (32%), Asian (48%), Multiracial (31%), Pacific Islander (38%), and White (36%) students and those enrolled in Average Needs (31%) and Low Needs (46%) districts.

Table 46D. ELA Grade 6 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	202,022	27.17	44.83	15.12	12.89	28.01
Gender	Female	99,570	22.65	45.72	16.40	15.23	31.63
	Male	102,452	31.56	43.96	13.87	10.62	24.49
Ethnicity	Asian	17,002	13.64	38.18	21.02	27.16	48.18
	Black	37,396	39.76	45.44	9.47	5.33	14.80
	Hispanic	48,322	36.51	47.44	10.37	5.68	16.05
	American Indian	1,127	36.56	44.81	11.00	7.63	18.63
	Multiracial	2,377	25.83	43.21	15.36	15.61	30.96
	Pacific Islander	408	22.06	40.20	19.12	18.63	37.75
	White	95,390	19.86	44.50	18.71	16.93	35.64
NRC	New York	65,808	30.44	44.19	13.22	12.16	25.38
	Big 4 Cities	7,542	54.44	36.17	6.01	3.38	9.39
	Urban/Suburban	14,334	40.46	44.62	9.64	5.27	14.92
	Rural	10,677	31.85	46.15	13.50	8.50	22.00
	Average Needs	53,490	22.83	46.29	16.94	13.94	30.88
	Low Needs	26,508	11.61	42.21	22.34	23.83	46.18
	Charter	8,182	28.15	51.66	13.58	6.61	20.19
	Non-Public	15,273	25.43	46.91	15.98	11.69	27.66
SWD	All Codes	31,913	67.30	28.60	2.91	1.19	4.11
SUA	All Codes	25,578	61.89	32.41	4.13	1.56	5.70
ELL	ELL=Y	16,447	71.44	25.71	2.15	0.70	2.85
SWD/SUA	SUA=504 plan codes	3,987	74.52	23.78	1.48	0.23	1.71
ELL/SUA	SUA & ELL codes	13,435	71.79	26.35	1.50	0.36	1.86

ELA Grade 7

Performance level distributions and n-counts of demographic subgroups for ELA Grade 7 are presented in Table 46E. Statewide, a combined 28% of students achieved Level III and Level IV. About 34% of Female students were at Level III or above, as compared to 23% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 50% of Asian students and 46% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 10–20% of students in those same performance categories. Only 3% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (28%): Female (34%), Asian (50%), Multiracial (30%), Pacific Islander (39%), and White (36%) students and those from Average Needs (30%) and Low Needs (46%) districts.

Table 46E. ELA Grade 7 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	199,911	35.33	36.26	22.63	5.78	28.41
Gender	Female	97,876	28.81	37.62	26.16	7.41	33.57
	Male	102,035	41.59	34.95	19.24	4.23	23.46
Ethnicity	Asian	16,553	18.67	30.98	35.57	14.78	50.35
	Black	37,363	49.80	35.16	13.36	1.69	15.04
	Hispanic	46,881	44.83	38.25	14.92	2.00	16.92
	American Indian	1,115	45.29	35.52	16.50	2.69	19.19
	Multiracial	2,084	38.53	31.91	21.50	8.06	29.56
	Pacific Islander	389	25.19	35.73	29.05	10.03	39.07
	White	95,526	27.75	36.73	27.86	7.65	35.51
NRC	New York	66,827	37.06	36.05	20.87	6.02	26.89
	Big 4 Cities	7,587	64.51	25.21	8.87	1.41	10.28
	Urban/Suburban	14,432	52.58	33.88	11.80	1.74	13.54
	Rural	11,223	44.23	35.91	16.95	2.91	19.86
	Average Needs	54,002	32.16	37.58	24.56	5.70	30.26
	Low Needs	28,278	17.20	36.34	35.36	11.10	46.46
	Charter	6,459	36.77	42.85	18.52	1.86	20.37
	Non-Public	10,849	34.02	38.15	23.17	4.65	27.83
SWD	All Codes	31,848	76.15	19.99	3.50	0.36	3.86
SUA	All Codes	23,910	73.70	21.38	4.37	0.55	4.92
ELL	ELL=Y	16,257	80.92	16.50	2.34	0.23	2.58
SWD/SUA	SUA=504 plan codes	3,072	88.09	11.04	0.81	0.07	0.88
ELL/SUA	SUA & ELL codes	11,017	83.43	15.11	1.33	0.13	1.46

ELA Grade 8

Performance level distributions and n-counts of demographic subgroups for ELA Grade 8 are presented in Table 46F. Statewide, a combined 35% of students achieved Level III and Level IV. About 40% of Female students were at Level III or above, as compared to 29% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 54% of Asian students and 56% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 13–27% of students in those same performance categories. Only 5% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (35%): Female (40%), Asian (54%), Multiracial (40%), Pacific Islander (45%), and White (44%) students and those enrolled in Average Needs (38%) and Low Needs (56%) districts and Non-Public (38%) schools.

Table 46F. ELA Grade 8 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	206,140	27.46	37.93	24.18	10.42	34.60
Gender	Female	100,645	22.18	37.78	26.94	13.10	40.04
	Male	105,495	32.50	38.09	21.56	7.86	29.42
Ethnicity	Asian	17,913	15.68	30.49	32.23	21.61	53.84
	Black	38,582	40.76	40.23	15.25	3.76	19.01
	Hispanic	47,589	36.55	42.03	16.92	4.49	21.42
	American Indian	1,096	34.95	39.96	19.71	5.38	25.09
	Multiracial	2,001	25.74	34.58	26.79	12.89	39.68
	Pacific Islander	438	22.60	32.88	25.11	19.41	44.52
	White	98,521	19.98	36.48	29.72	13.82	43.54
NRC	New York	69,167	31.55	39.47	20.55	8.44	28.99
	Big 4 Cities	7,330	56.94	30.12	10.19	2.74	12.93
	Urban/Suburban	14,111	40.54	39.35	15.75	4.37	20.11
	Rural	11,029	33.91	38.80	20.44	6.85	27.29
	Average Needs	54,214	23.36	38.16	26.84	11.63	38.47
	Low Needs	28,946	11.70	32.73	35.69	19.88	55.57
	Charter	5,067	27.25	46.54	21.59	4.62	26.21
	Non-Public	16,028	22.50	39.03	27.46	11.01	38.46
SWD	All Codes	30,866	68.07	26.67	4.59	0.66	5.26
SUA	All Codes	24,518	63.99	28.51	6.35	1.15	7.51
ELL	ELL=Y	16,442	71.26	24.30	4.01	0.43	4.44
SWD/SUA	SUA=504 plan codes	3,149	85.04	13.88	1.05	0.03	1.08
ELL/SUA	SUA & ELL codes	10,832	81.03	17.22	1.66	0.09	1.75

Mathematics Test Performance Level Distributions

Table 47 shows the performance level distributions for all examinees from public, charter, and private schools with valid scores, and presents mathematics performance level data for total populations of students in Grades 3–8. Performance level data for selected subgroups of students were also examined. In general, these summaries reflect the same achievement trends as in the scale score summary discussion. Male and Female students performed similarly across grades. More White, Native Hawaiian or Other Pacific Islander, and Asian students were classified in Level III and above, as compared to their peers from other ethnic subgroups. Students from Low and Average Needs districts outperformed students from High Needs districts (New York City, Big 4 Cities, High Needs Urban/Suburban, and High Needs Rural), Private Schools, and Charter Schools. The subgroups that used the Korean or Chinese translations outperformed other test translation subgroups. The Level III and above rates for SWD and SUA subgroups were low compared to the total population of examinees. Please note that the case counts for the Haitian-Creole, Korean, and Russian translation subgroups were very low, and the results might have been heavily influenced by very high and/or very low achieving individual students.

Table 47. Mathematics Test Performance Level Distributions

Grade	N-Count	Level I	Level II	Level III	Level IV	Level III & IV
3	201,357	27.02	31.07	26.15	15.76	41.91
4	206,007	26.95	31.92	23.91	17.23	41.14
5	200,573	32.49	28.75	24.00	14.77	38.76
6	200,645	27.43	35.82	19.40	17.35	36.75
7	195,382	35.47	32.85	22.36	9.32	31.68
8	161,756	36.88	41.17	17.10	4.85	21.95

Mathematics Grade 3

Performance level summaries and n-counts of demographic subgroups for Grade 3 are presented in Table 48A. Statewide, a combined 42% of students achieved Level III and Level IV. About 42% of both Female and Male students were at Level III or above. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 66% of Asian students and 63% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 17–35% of students in those same performance categories. Only 17% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (42%): Asian (66%), Multiracial (44%), Pacific Islander (47%), and White (51%) students and those from Average Needs (47%) and Low Needs (63%) districts and Charter (50%) schools. For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 2% (Haitian-Creole) to 63% (Korean).

Table 48A. Mathematics Grade 3 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	201,357	27.02	31.07	26.15	15.76	41.91
Gender	Female	98,417	26.11	32.29	26.55	15.04	41.60
	Male	102,940	27.89	29.90	25.76	16.45	42.21
Ethnicity	Asian	17,717	10.81	22.87	32.80	33.52	66.33
	Black	35,789	41.99	32.82	17.83	7.36	25.19
	Hispanic	52,030	36.53	34.83	20.80	7.84	28.65
	American Indian	1,172	29.78	33.11	24.66	12.46	37.12
	Multiracial	3,630	26.25	29.67	24.60	19.48	44.08
	Pacific Islander	520	23.27	30.00	28.65	18.08	46.73
	White	90,499	18.83	29.85	31.28	20.04	51.32
NRC	New York	72,565	29.90	31.40	23.89	14.81	38.71
	Big 4 Cities	7,956	55.54	27.80	12.04	4.61	16.65
	Urban/Suburban	15,313	43.09	33.29	17.76	5.86	23.62
	Rural	10,407	30.78	34.28	24.42	10.53	34.95
	Average Needs	51,437	21.45	31.70	29.61	17.25	46.86
	Low Needs	24,324	11.16	25.51	35.56	27.77	63.33
	Charter	8,332	18.00	32.43	29.20	20.37	49.57
	Non-Public	10,807	29.51	33.47	25.06	11.96	37.02
SWD	All Codes	29,981	57.61	26.09	11.74	4.56	16.30
SUA	All Codes	29,116	50.79	29.01	14.77	5.44	20.20
ELL	ELL=Y	15,349	61.65	24.48	10.34	3.52	13.86
SWD/SUA	SUA=504 plan codes	8,021	54.29	29.22	12.39	4.09	16.48
ELL/SUA	SUA & ELL codes	17,810	55.08	28.87	12.20	3.85	16.05
ELL Test Language	English	14,434	54.45	29.52	12.33	3.70	16.02
	Chinese	571	13.84	34.15	32.40	19.61	52.01
	Haitian-Creole	51	80.39	17.65	1.96	0.00	1.96
	Korean	35	5.71	31.43	40.00	22.86	62.86
	Russian	84	39.29	30.95	20.24	9.52	29.76
	Spanish	2,635	68.12	24.29	6.68	0.91	7.59
	All Translations	3,376	57.76	26.10	11.64	4.50	16.14

Mathematics Grade 4

Performance level summaries and n-counts of demographic subgroups for Grade 4 are presented in Table 48B. Statewide, a combined 41% of students achieved Level III and Level IV. About 41% of both Female and Male students were at Level III or above. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 69% of Asian students and 61% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 17–31% of students in those same performance categories. Only 13% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher

percentage of students in Levels III and IV than statewide (41%): Asian (69%), Multiracial (44%), Pacific Islander (46%), and White (49%) students and those enrolled in Average Needs (45%) and Low Needs (61%) districts and Charter (50%) schools. For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 4% (Haitian-Creole) to 66% (Korean).

Table 48B. Mathematics Grade 4 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	206,007	26.95	31.92	23.91	17.23	41.14
Gender	Female	101,235	26.15	32.94	24.34	16.57	40.91
	Male	104,772	27.72	30.93	23.49	17.86	41.35
Ethnicity	Asian	18,471	9.84	20.98	28.37	40.81	69.18
	Black	36,054	41.67	32.96	16.74	8.63	25.36
	Hispanic	50,470	37.49	34.82	18.61	9.08	27.70
	American Indian	1,136	32.13	34.51	19.81	13.56	33.36
	Multiracial	2,997	25.96	29.73	25.16	19.15	44.31
	Pacific Islander	468	22.01	31.62	25.85	20.51	46.37
	White	96,411	19.19	32.14	28.51	20.16	48.67
NRC	New York	70,616	29.27	30.64	21.35	18.73	40.09
	Big 4 Cities	7,929	56.10	27.14	12.13	4.63	16.76
	Urban/Suburban	14,778	44.28	32.84	16.25	6.62	22.88
	Rural	10,453	33.59	35.23	21.58	9.60	31.18
	Average Needs	52,444	21.61	33.66	27.37	17.36	44.73
	Low Needs	26,002	10.66	27.95	33.14	28.25	61.39
	Charter	7,066	17.17	32.59	27.36	22.88	50.24
	Non-Public	16,507	30.08	37.01	21.76	11.15	32.91
SWD	All Codes	31,445	60.73	26.37	9.15	3.74	12.89
SUA	All Codes	32,190	55.68	29.00	11.06	4.26	15.31
ELL	ELL=Y	18,651	64.66	25.33	7.64	2.37	10.01
SWD/SUA	SUA=504 plan codes	8,162	61.63	26.64	8.28	3.46	11.74
ELL/SUA	SUA & ELL codes	18,426	57.32	27.99	10.25	4.44	14.70
ELL Test Language	English	14,951	56.49	28.97	10.27	4.27	14.53
	Chinese	569	14.06	29.00	33.22	23.73	56.94
	Haitian-Creole	45	80.00	15.56	4.44	0.00	4.44
	Korean	32	3.13	31.25	28.13	37.50	65.63
	Russian	110	40.91	36.36	13.64	9.09	22.73
	Spanish	2,719	71.83	22.18	5.11	0.88	5.99
	All Translations	3,475	60.86	23.74	10.19	5.21	15.40

Mathematics Grade 5

Performance level summaries and n-counts of demographic subgroups for Grade 5 are presented in Table 48C. Statewide, a combined 39% of students achieved Level III and Level IV. About 39% of both Female and Male students were at Level III or above. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 69% of Asian students and 61% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 14–27% of students in those same performance categories. Only 11% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (39%): Asian (69%), Pacific Islander (45%), and White (46%) students and those from Average Needs (41%) and Low Needs (61%) districts. For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 3% (Spanish) to 57% (Korean).

Table 48C. Mathematics Grade 5 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	200,573	32.49	28.75	24.00	14.77	38.76
Gender	Female	98,329	30.93	30.07	24.71	14.29	39.00
	Male	102,244	34.00	27.48	23.31	15.22	38.53
Ethnicity	Asian	17,680	11.97	19.51	30.18	38.33	68.51
	Black	35,684	48.65	29.60	15.70	6.04	21.75
	Hispanic	48,900	42.61	31.50	18.80	7.10	25.90
	American Indian	1,134	37.21	31.13	19.93	11.73	31.66
	Multiracial	2,578	34.64	27.15	22.85	15.36	38.21
	Pacific Islander	466	26.82	28.54	28.33	16.31	44.64
	White	94,131	24.87	28.75	28.73	17.64	46.38
NRC	New York	68,577	33.02	28.15	22.61	16.22	38.83
	Big 4 Cities	7,754	64.11	22.09	9.97	3.83	13.80
	Urban/Suburban	14,641	52.54	28.04	14.83	4.58	19.41
	Rural	10,455	41.45	31.58	19.81	7.15	26.96
	Average Needs	53,329	28.34	30.73	26.64	14.29	40.93
	Low Needs	26,388	13.33	25.81	34.11	26.75	60.86
	Charter	8,137	30.63	32.63	24.67	12.08	36.75
	Non-Public	11,059	39.10	29.95	21.07	9.88	30.95
SWD	All Codes	32,072	67.91	21.57	7.97	2.55	10.52
SUA	All Codes	29,176	63.85	23.30	9.86	2.99	12.85
ELL	ELL=Y	17,483	72.51	19.40	6.57	1.52	8.09
SWD/SUA	SUA=504 plan codes	6,204	71.32	20.16	6.21	2.30	8.51
ELL/SUA	SUA & ELL codes	15,257	66.72	22.04	7.99	3.25	11.24

Table 48C. Mathematics Grade 5 Performance Level Distribution by Subgroup (cont.)

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
ELL Test Language	English	11,931	66.43	22.82	7.90	2.84	10.75
	Chinese	530	15.47	28.68	30.19	25.66	55.85
	Haitian-Creole	49	85.71	8.16	6.12	0.00	6.12
	Korean	35	11.43	31.43	28.57	28.57	57.14
	Russian	81	43.21	25.93	24.69	6.17	30.86
	Spanish	2,631	79.44	17.18	3.15	0.23	3.38
	All Translations	3,326	67.74	19.24	8.30	4.72	13.02

Mathematics Grade 6

Performance level summaries and n-counts of demographic subgroups for Grade 6 are presented in Table 48D. Statewide, a combined 37% of students achieved Level III and Level IV. About 38% of Female students were at Level III or above, as compared to 36% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 67% of Asian students and 62% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 13–26% of students in those same performance categories. Only 10% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (37%): Asian (67%), Pacific Islander (42%), and White (45%) students and those enrolled in Average Needs (40%) and Low Needs (62%) districts and Charter (40%) schools. For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 1% (Haitian-Creole) to 75% (Korean).

Table 48D. Mathematics Grade 6 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	200,645	27.43	35.82	19.40	17.35	36.75
Gender	Female	98,538	25.35	36.93	20.27	17.44	37.72
	Male	102,107	29.44	34.75	18.56	17.25	35.81
Ethnicity	Asian	17,426	10.00	22.87	22.90	44.23	67.14
	Black	37,231	43.14	37.03	12.86	6.97	19.83
	Hispanic	49,019	38.41	38.61	14.81	8.17	22.98
	American Indian	1,095	36.35	38.72	14.52	10.41	24.93
	Multiracial	2,319	27.21	35.23	19.92	17.64	37.56
	Pacific Islander	433	23.79	33.95	19.86	22.40	42.26
	White	93,122	18.56	36.28	23.81	21.35	45.16

Table 48D. Mathematics Grade 6 Performance Level Distribution by Subgroup (cont.)

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
NRC	New York	67,002	32.36	33.74	16.57	17.33	33.90
	Big 4 Cities	7,628	56.36	30.73	8.64	4.27	12.91
	Urban/Suburban	14,305	43.25	37.60	13.26	5.89	19.15
	Rural	10,431	31.72	41.99	17.52	8.77	26.29
	Average Needs	51,565	21.22	38.95	22.48	17.35	39.83
	Low Needs	25,929	9.33	29.17	27.38	34.12	61.50
	Charter	8,161	20.78	39.44	22.52	17.25	39.77
	Non-Public	15,434	28.81	40.18	18.70	12.31	31.01
SWD	All Codes	31,225	64.61	26.91	5.87	2.61	8.48
SUA	All Codes	28,565	57.02	30.58	8.37	4.04	12.41
ELL	ELL=Y	17,488	65.75	26.81	5.59	1.85	7.44
SWD/SUA	SUA=504 plan codes	5,755	64.78	25.14	5.98	4.10	10.08
ELL/SUA	SUA & ELL codes	15,360	63.31	26.75	6.33	3.61	9.94
ELL Test Language	English	11,130	63.09	27.61	6.24	3.05	9.30
	Chinese	600	11.33	34.33	26.67	27.67	54.33
	Haitian-Creole	140	81.43	17.86	0.71	0.00	0.71
	Korean	36	8.33	16.67	22.22	52.78	75.00
	Russian	119	45.38	34.45	8.40	11.76	20.17
	Spanish	3,335	73.85	22.73	2.94	0.48	3.42
	All Translations	4,230	63.88	24.49	6.55	5.08	11.63

Mathematics Grade 7

Performance level summaries and n-counts of demographic subgroups for Grade 7 are presented in Table 48E. Statewide, a combined 32% of students achieved Level III and Level IV. About 32% of Female students were at Level III or above, as compared to 31% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 62% of Asian students and 54% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 9–21% of students in those same performance categories. Only 6% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (32%): Asian (62%), Pacific Islander (44%), and White (40%) students and those from Average Needs (35%) and Low Needs (54%) districts. For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 1% (Haitian-Creole) to 58% (Korean).

Table 48E. Mathematics Grade 7 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	195,382	35.47	32.85	22.36	9.32	31.68
Gender	Female	95,301	33.44	34.12	23.21	9.24	32.44
	Male	100,081	37.41	31.64	21.55	9.40	30.95
Ethnicity	Asian	16,746	14.22	23.83	33.17	28.78	61.96
	Black	36,907	54.86	30.70	11.91	2.52	14.43
	Hispanic	47,146	48.25	33.95	14.65	3.15	17.80
	American Indian	1,079	45.78	33.18	15.94	5.10	21.04
	Multiracial	1,976	36.99	30.62	21.66	10.73	32.39
	Pacific Islander	392	23.98	32.14	27.04	16.84	43.88
	White	91,136	24.81	34.85	28.67	11.68	40.34
NRC	New York	67,826	39.69	30.59	19.29	10.43	29.72
	Big 4 Cities	7,575	69.16	21.76	7.13	1.95	9.08
	Urban/Suburban	14,240	55.84	30.70	11.48	1.99	13.47
	Rural	10,784	41.35	37.74	17.70	3.21	20.91
	Average Needs	50,792	28.62	36.77	26.03	8.58	34.61
	Low Needs	26,588	14.77	30.75	35.96	18.52	54.48
	Charter	6,427	31.62	38.20	24.16	6.02	30.19
	Non-Public	10,910	37.75	36.28	19.69	6.28	25.97
SWD	All Codes	30,757	74.91	19.66	4.62	0.81	5.43
SUA	All Codes	24,654	69.33	22.70	6.46	1.51	7.97
ELL	ELL=Y	15,926	76.56	18.76	4.04	0.64	4.68
SWD/SUA	SUA=504 plan codes	4,271	78.79	14.80	4.59	1.83	6.42
ELL/SUA	SUA & ELL codes	12,962	74.14	18.24	5.85	1.77	7.62
ELL Test Language	English	8,821	75.03	18.81	4.87	1.29	6.17
	Chinese	647	15.92	29.37	38.18	16.54	54.71
	Haitian-Creole	141	85.82	13.48	0.71	0.00	0.71
	Korean	40	10.00	32.50	37.50	20.00	57.50
	Russian	105	55.24	34.29	10.48	0.00	10.48
	Spanish	3,208	84.35	13.93	1.68	0.03	1.71
	All Translations	4,141	72.25	17.02	7.92	2.80	10.72

Mathematics Grade 8

It should be noted that students in the 2013 population did not have available to them the double testing waiver and hence included more students who were accelerated and taking the Integrated Algebra Regents Examination. So the 2014 population—summarized in Table 48F—was smaller (n-count = 161,756) than the 2013 population (n-count = 214,932) and had slightly lower percentages of students earning Level III or above: 21.95% in 2014 as opposed to 26.91% in 2013.

Performance level summaries and n-counts of demographic subgroups for Grade 8 are presented in F. Statewide, a combined 22% of students achieved Level III and Level IV. About 23% of

Female students were at Level III or above, as compared to 21% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 51% of Asian students and 38% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 7–15% of students in those same performance categories. Only 6% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (22%): Asian (51%), Pacific Islander (30%), and White (26%) students and those enrolled in Low Needs (38%) districts and Charter (35%) and Non-Public (25%) schools. For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 0% (Haitian-Creole) to 56% (Korean).

Table 48F. Mathematics Grade 8 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	161,756	36.88	41.17	17.10	4.85	21.95
Gender	Female	77,575	35.05	42.04	17.83	5.08	22.90
	Male	84,181	38.56	40.37	16.42	4.64	21.07
Ethnicity	Asian	12,099	16.76	32.50	29.73	21.01	50.74
	Black	34,585	52.36	35.60	9.85	2.19	12.04
	Hispanic	43,081	45.50	39.57	12.41	2.52	14.93
	American Indian	909	48.07	38.28	11.88	1.76	13.64
	Multiracial	1,494	36.48	41.97	17.07	4.48	21.55
	Pacific Islander	325	33.54	36.92	22.46	7.08	29.54
	White	69,263	27.18	46.51	21.47	4.84	26.31
NRC	New York	59,604	39.39	37.75	16.08	6.78	22.86
	Big 4 Cities	6,967	70.32	22.75	5.73	1.21	6.93
	Urban/Suburban	12,408	55.92	36.67	6.75	0.66	7.41
	Rural	9,106	42.48	45.78	10.71	1.03	11.74
	Average Needs	37,396	30.62	48.71	18.35	2.32	20.67
	Low Needs	16,124	16.97	45.04	30.53	7.45	37.99
	Charter	4,870	25.56	39.63	25.61	9.20	34.80
	Non-Public	15,110	32.59	41.95	18.69	6.78	25.47
SWD	All Codes	28,623	70.65	24.45	4.34	0.55	4.89
SUA	All Codes	20,867	66.21	27.07	5.83	0.89	6.72
ELL	ELL=Y	13,887	71.83	23.90	3.97	0.30	4.27
SWD/SUA	SUA=504 plan codes	3,215	78.79	15.77	3.86	1.59	5.44
ELL/SUA	SUA & ELL codes	12,646	67.74	23.81	6.25	2.20	8.45
ELL Test Language	English	8,445	69.56	23.60	5.39	1.46	6.84
	Chinese	694	14.70	33.14	32.85	19.31	52.16
	Haitian-Creole	183	81.42	18.58	0.00	0.00	0.00
	Korean	34	17.65	26.47	41.18	14.71	55.88
	Russian	114	43.86	35.09	14.91	6.14	21.05
	Spanish	3,176	75.13	22.20	2.39	0.28	2.68
	All Translations	4,201	64.10	24.23	7.97	3.69	11.66

References

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.
- Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika* 37:29–51.
- Bock, R.D. & M. Aitkin (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika* 46:443–459.
- Cattell, R.B. (1966). The Screen Test for the Number of Factors. *Multivariate Behavioral Research* 1:245–276.
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16:297–334.
- Dorans, N.J., A.P. Schmitt & C.A. Bleistein (1992). The standardization approach to assessing comprehensive differential item functioning. *Journal of Educational Measurement* 29:309–319.
- Dorans, N.J. & P. W. Holland (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35–66). Hillsdale, NJ: Lawrence Erlbaum.
- Fleiss J.L. & J. Cohen (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33: 613–619.
- Green, D.R., W.M. Yen & G.R. Burket (1989). Experiences in the application of item response theory in test construction. *Applied Measurement in Education* 2:297–312.
- Huynh, H. & C. Schneider (2004). *Vertically moderated standards as an alternative to vertical scaling: assumptions, practices, and an odyssey through NAEP*. Paper presented at the National Conference on Large-Scale Assessment. Boston, MA, June 21.
- Jensen, A.R. (1980). *Bias in mental testing*. New York: Free Press.
- Johnson, N.L. & S. Kotz (1970). *Distributions in Statistics: Continuous Univariate Distributions*, Vol. 2. New York: John Wiley.
- Kim, S. & M. J. Kolen (2004). *STUIRT: A computer program for scale transformation under unidimensional item response theory models*. Iowa City, IA: Iowa Testing Programs, The University of Iowa.
- Kolen, M.J. & Z. Cui (2004). *POLYEQUATE*. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.
- Kolen, M.J. & R.L. Brennan (1995). *Test Equating: Methods and Practices*. New York: Springer-Verlag.
- Landis, J. R. & G. G. Koch. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174.

- Lee, W. C., B.A. Hanson & R.L. Brennan (2002). Estimating consistency and accuracy indices for multiple classifications. *Applied Psychological Measurement* 26:412–432.
- Lee, W. C. (2008). *Classification consistency and accuracy for complex assessments using item response theory*. (CASMA Research Report No. 27). Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.
- Lee, W. C. & M. J. Kolen (2006, Revised 2008). IRT-CLASS (Version 2.0). Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.
- Linn, R.L. (1991). Linking results of distinct assessments. *Applied Measurement in Education* 6(1): 83–102.
- Linn, R.L. & D. Harnisch (1981). Interactions between item content and group membership on achievement test items. *Journal of Educational Measurement* 18: 109–118.
- Livingston, S.A. & C. Lewis (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement* 32: 179–197.
- Lord, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum.
- Lord, F.M. & M.R. Novick (1968). *Statistical Theories of Mental Test Scores*. Menlo Park, CA: Addison-Wesley.
- Mehrens, W.A. & I.J. Lehmann (1991). *Measurement and Evaluation in Education and Psychology*, 3rd ed. New York: Holt, Rinehart, and Winston.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement* 16: 159–176.
- Muraki, E. & R.D. Bock (1991). *PARSCALE: Parameter Scaling of Rating Data* [Computer program]. Chicago, IL: Scientific Software, Inc.
- Novick, M.R. & P.H. Jackson (1974). *Statistical Methods for Educational and Psychological Research*. New York: McGraw-Hill.
- NYSED. (2014) New York State Testing Program 2013: English Language Arts and Mathematics Grades 3–8 Technical Report. Albany, NY: New York State Education Department (NYSED). Retrieved from: <http://www.p12.nysed.gov/assessment/reports/2013/ela-math-tr13.pdf>
- Qualls, A.L. (1995). Estimating the reliability of a test containing multiple-item formats. *Applied Measurement in Education* 8: 111–120.
- Reckase, M.D. (1979). Unifactor latent trait models applied to multifactor tests: results and implications. *Journal of Educational Statistics* 4: 207–230.
- Sandoval, J.H. & M.P. Mille (1979) *Accuracy of judgments of WISC-R item difficulty for minority groups*. Paper presented at the annual meeting of the American Psychological Association, New York. August.
- Stocking, M.L. & F.M. Lord (1983). Developing a common metric in item response theory. *Applied Psychological Measurement* 7: 201–210.

- Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika* 47: 175–186.
- Thissen, D. (1991). *MULTILOG* [Computer program]. Chicago, IL: Scientific Software, Inc.
- Thompson, S.J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal Design Applied to Large Scale Assessments (NCEO Synthesis Report 44)*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from: <http://www.cehd.umn.edu/nceo/onlinepubs/Synthesis44.html>.
- Wang, T.M., J. Kolen, & D.J. Harris (2000). Psychometric properties of scale scores and performance levels for performance assessment using polytomous IRT. *Journal of Educational Measurement* 37: 141–162.
- Yen, W.M. (1997). The technical quality of performance assessments: Standard errors of percents of students reaching standards. *Educational Measurement: Issues and Practice*: 5–15.
- Yen, W.M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement* 30: 187–213.
- Yen, W. M. (1984). Obtaining maximum likelihood trait estimates from number correct scores for the three-parameter logistic model. *Journal of Educational Measurement* 21:93–111.
- Yen, W.M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement* 5: 245–262.
- Yen, W.M., R.C. Sykes, K. Ito & M. Julian (1997). *A Bayesian/IRT index of objective performance for tests with mixed-item types*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago: March.
- Zwick, R., J.R. Donoghue & A. Grima, (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement* 36: 225–33.

Appendix A: ELA and Mathematics Test Configuration and Testing Times

Table A1. ELA Test Configuration

Grade	Day	Book	Number of Items				
			Multiple-Choice		Constructed-Response		Total
			Operational	Embedded	Operational	Embedded	
3*	1	1	24	6	0	0	30
	2	2	7	0	4	0	11
	3	3	0	0	6	0	6
	Total		31	6	10	0	47
4	1	1	24	6	0	0	30
	2	2	7	0	4	0	11
	3	3	0	0	6	0	6
	Total		31	6	10	0	47
5	1	1	35	7	0	0	42
	2	2	7	0	4	0	11
	3	3	0	0	6	0	6
	Total		42	7	10	0	59
6	1	1	35	7	0	0	42
	2	2	7	0	4	0	11
	3	3	0	0	6	0	6
	Total		42	7	10	0	59
7**	1	1	35	7	0	0	42
	2	2	7	0	4	0	11
	3	3	0	0	6	0	6
	Total		42	7	10	0	59
8	1	1	35	7	0	0	42
	2	2	7	0	4	0	11
	3	3	0	0	6	0	6
	Total		42	7	10	0	59

*For Grade 3, three items were excluded from the analysis and scoring because of greater than expected omit rates.

**For Grade 7, one item was excluded from the analysis and scoring because of poor fit to the item response theory (IRT) model.

Table A2. Mathematics Test Configuration

Grade	Day	Book	Number of Items				
			Multiple-Choice		Constructed-Response		Total
			Operational	Embedded	Operational	Embedded	
3	1	1	20	4	0	0	24
	2	2	21	3	0	0	24
	3	3	0	0	8	0	8
	Total		41	7	8	0	56
4	1	1	20	4	0	0	24
	2	2	22	3	0	0	25
	3	3	0	0	10	0	10
	Total		42	7	10	0	59
5	1	1	20	4	0	0	24
	2	2	22	3	0	0	25
	3	3	0	0	10	0	10
	Total		42	7	10	0	59
6	1	1	24	4	0	0	28
	2	2	24	3	0	0	27
	3	3	0	0	10	0	10
	Total		48	7	10	0	65
7	1	1	24	4	0	0	28
	2	2	24	3	0	0	27
	3	3	0	0	10	0	10
	Total		48	7	10	0	65
8	1	1	24	4	0	0	28
	2	2	24	3	0	0	27
	3	3	0	0	10	0	10
	Total		48	7	10	0	65

Table A3. ELA Testing Times

Grade	Day	Book	Estimated Time on Task (min)	Session Time (min)
3	1	1	50	70
	2	2	50	70
	3	3	50	70
	Total		150	210
4	1	1	50	70
	2	2	50	70
	3	3	50	70
	Total		150	210
5	1	1	70	90
	2	2	60	90
	3	3	50	90
	Total		180	270
6	1	1	70	90
	2	2	60	90
	3	3	50	90
	Total		180	270
7	1	1	70	90
	2	2	60	90
	3	3	50	90
	Total		180	270
8	1	1	70	90
	2	2	60	90
	3	3	50	90
	Total		180	270

Source: 2014 *Common Core ELA and Mathematics Test Guides*.

The ELA estimated times on task were based on the following rules of thumb:

- Average time to read a passage—5 minutes;
- Average time to respond to a multiple-choice question—1 minute;
- Average time to respond to a two-point constructed response question—3 minutes; and
- Average time to responds to a four-point constructed response question—20 minutes.

Table A4. Math Testing Times

Grade	Day	Book	Estimated Time on Task (min)	Session Time (min)
3	1	1	40	60
	2	2	40	60
	3	3	50	70
	Total		130	190
4	1	1	40	60
	2	2	40	60
	3	3	70	90
	Total		150	210
5	1	1	40	80
	2	2	40	80
	3	3	70	90
	Total		150	250
6	1	1	40	80
	2	2	40	80
	3	3	70	90
	Total		150	250
7	1	1	40	80
	2	2	40	80
	3	3	70	90
	Total		150	250
8	1	1	40	80
	2	2	40	80
	3	3	70	90
	Total		150	250

Source: 2014 *Common Core ELA and Mathematics Test Guides*.

The mathematics estimated times on task were based on the following rules of thumb:

- Average time to respond to a multiple-choice question—1.5 minutes;
- Average time to respond to a two-point constructed response question—5 minutes; and
- Average time to responds to a three-point constructed response question—9 minutes.

The testing times listed above do not include approximately 10 minutes reserved for preparation at the beginning of each session for handing out materials and reading directions. Additional details on security, scheduling, classroom organization and preparation, test materials, and administration can be found in the 2014 *Teacher's Directions* and the *School Administrator's Manual*, which are accessible online:

- 2014 *Common Core ELA Teacher's Directions*
 - Grades 3–5: <http://www.p12.nysed.gov/assessment/sam/ei/td-35ela14.pdf>
 - Grades 6–8: <http://www.p12.nysed.gov/assessment/sam/ei/td-68ela14.pdf>

- *2014 Common Core Mathematics Teacher’s Directions*
- Grades 3–5: <http://www.p12.nysed.gov/assessment/sam/ei/td-35math14.pdf>
- Grades 6–8: <http://www.p12.nysed.gov/assessment/sam/ei/td-68math14.pdf>
- *2014 Common Core ELA and Mathematics Tests School Administrator’s Manual*
 - <http://www.p12.nysed.gov/assessment/sam/ei/eisam14rev.pdf>
- *2014 Common Core ELA and Mathematics Test Guides*
 - <https://www.engageny.org/resource/test-guides-for-english-language-arts-and-mathematics>

Appendix B: ELA and Mathematics Test Blueprints

Table B1. ELA Test Blueprint

Grade	Total Points on OP Test	Standard	Point Range		% of Test	
			Target	Actual	Target	Actual
3	55	Literature	18–44	22	33–80%	45%
		Information	18–44	26	33–80%	53%
		Language	1–4	1	2–7%	2%
4	55	Literature	18–44	27	33–80%	49%
		Information	18–44	26	33–80%	47%
		Language	1–4	2	2–7%	4%
5	66	Literature	18–51	30	27–77%	45%
		Information	18–51	35	27–77%	53%
		Language	1–4	1	2–6%	2%
6	65	Literature	11–44	22	17–67%	33%
		Information	25–58	42	38–88%	64%
		Language	1–4	2	2–6%	3%
7	66	Literature	11–44	25	17–67%	38%
		Information	25–58	39	38–88%	60%
		Language	1–4	1	2–6%	2%
8	66	Literature	11–44	28	17–67%	42%
		Information	25–58	36	38–88%	55%
		Language	1–4	2	2–6%	3%

Table B2. Mathematics Test Blueprint

Grade	Total Points on OP Test	Standard	Point Range		% of Test	
			Target	Actual	Target	Actual
3	60	Operations and Algebraic Thinking	23–31	27	38–52%	45%
		Number and Operations in Base Ten	3–5	4	5–8%	7%
		Number and Operations – Fractions	10–14	12	17–23%	20%
		Measurement and Data	12–18	15	20–30%	25%
		Geometry*	1–3	2	2–5%	3%
4	66	Operations and Algebraic Thinking	11–15	13	17–23%	20%
		Number and Operations in Base Ten	14–20	17	21–30%	26%
		Number and Operations – Fractions	15–21	18	23–32%	27%
		Measurement and Data	9–15	12	14–23%	18%
		Geometry	5–7	6	8–11%	9%
5	66	Operations and Algebraic Thinking	3–5	4	5–8%	6%
		Number and Operations in Base Ten	15–21	18	23–32%	27%
		Number and Operations – Fractions	22–28	25	33–42%	38%
		Measurement and Data	13–21	17	20–32%	26%
		Geometry*	1–3	2	2–5%	3%
6	72	Ratios and Proportional Relationships	16–20	18	22–28%	25%
		The Number System	13–19	16	18–26%	22%
		Expressions and Equations	23–33	28	32–46%	39%
		Geometry	8–12	10	11–17%	14%
7	72	Ratios and Proportional Relationships	18–22	20	25–31%	28%
		The Number System	12–16	14	17–22%	19%
		Expressions and Equations	19–25	22	26–35%	31%
		Geometry	3–7	5	4–10%	7%
		Statistics and Probability	8–14	11	11–19%	15%
8	72	Expressions and Equations	26–34	30	36–47%	42%
		Functions	14–20	19	19–28%	26%
		Geometry	16–22	17	22–31%	24%
		Statistics and Probability	5–7	6	7–10%	8%

* There is a slight difference between the “Target% of Test” shown in these tables and the tables presented in the Guides to the 2014 Common Core Mathematics Tests. The guides were intended to provide general guidance regarding content coverage of mathematics domains so that classroom instruction would continue to cover the depth and breadth of the Common Core mathematics standards.

Appendix C: Passage Selection Guidelines for Assessing ELA

General Guidelines

Along with instructional materials and teacher training, assessment development is essential to the successful implementation of the CCSS. While many of the expectations outlined in the CCSS align with previous versions of the NYS Learning Standards for ELA, the CCSS do represent some shifts in emphasis with direct implications for assessment development. In particular the CCSS devote considerable attention to the types and nature of texts used in instruction and assessment. The foundation for preparing students for the linguistic rigors of college and the workplace lies in the texts with which they interact. By the time they graduate, students should be prepared to successfully read and analyze the types of complex texts they will encounter after high school. Selecting passages of appropriate type and complexity for use in assessment is integral to this preparation.

One of the major shifts of the CCSS is an emphasis on developing skills for comprehending and analyzing informational texts. Increased exposure to informational texts better prepares students for the various types of texts they will encounter in college and the workplace. The array of passages selected for assessment from K–12 should support the development of the necessary skills to handle this range of informational texts.

Another shift is an increased emphasis on the analysis across multiple texts, often of varied genres and media. Several standards, especially for reading literature, require intertextual and multi-media analysis. These expectations require special attention to the selection of related passages, chosen specifically to support the assessment of the full range of expectations. It will also require careful consideration of which standards are appropriate for large-scale assessment formats, and how these assessments might be modified to include passages of a variety of media.

In addition to the usual fairness and sensitivity guidelines when selecting passages for assessment, attention should be dedicated to three additional considerations:

Text Complexity

Text Types

Text Suitability for Specific Standards

These guidelines should inform the training of passage finders in order to ensure a pool of acceptable passages that can support assessment of all the CCSS Reading Informational Texts standards. They should also alert form assemblers as they construct forms that will assess the complete range of skills.

Appendix D: Universal Design Item Checklist

Universal Design Item Checklist	
A.	Precisely Designed Constructs
Definition	The item construct is clearly defined so that all irrelevant cognitive, sensory, emotional, and physical barriers are removed.
√	The item does not add skills to those being measured (no extraneous skills tested).
B.	Language Appropriateness
Definition	The item avoids words or phrases that are sexist, racist, or otherwise offensive, inappropriate, or negative to any subgroup. Language should be simple and clear.
√	The item uses commonly used words—simpler is better.
√	The item uses vocabulary appropriate for the grade level.
√	Idiomatic speech and figurative language are avoided unless being measured.
√	The item avoids technical terms unrelated to the content.
√	The item contains no unnecessary words.
√	The sentence complexity contained in the item is appropriate for the grade level.
√	The item avoids ambiguous or multiple-meaning words (e.g., crane—the bird—can easily be confused with crane—heavy machinery).
√	All pronouns have clear referents.
√	The item avoids the use of proper names. (Such names may be unfamiliar or difficult for cultural subgroups.)
√	The item avoids irregularly spelled words.
C.	Gender Stereotypes
Definition	The item avoids stereotyping as results of associating genders with certain professions or activities. All groups of society should be portrayed accurately and fairly regarding gender.
√	The item is free of content that might offend a gender subgroup.
√	The item is free of content that might unfairly advantage or disadvantage a gender subgroup.
D.	Ethnic Stereotypes
Definition	The item avoids unnecessary references to and uses the proper reference for ethnic, racial, or cultural groups.
√	The item is free of content that might offend an ethnic subgroup.
√	The item is free of content that might unfairly advantage or disadvantage an ethnic subgroup.
√	The artwork included in an item adequately reflects the diversity of the student population.
E.	Cultural Familiarity
Definition	Does not rely on an assumed shared experience that is class oriented or native English speaking oriented. Presentations of cultural or ethnic differences should neither explicitly nor implicitly rely on stereotypes nor make moral judgments.
√	The item does not rely on an assumed shared experience that is class oriented or native English speaking oriented.
√	The item is free from content that might offend a socioeconomic subgroup.
√	The item is free of content that might unfairly advantage or disadvantage a socioeconomic subgroup.

Universal Design Item Checklist	
√	The item is free from unnecessary cultural references.
√	The item is free from religious references.
F.	Geographic Bias
Definition	All groups of society should be portrayed accurately and fairly regarding geographic setting. A particular geographic setting shouldn't be used repeatedly, and urban, suburban, and rural settings should be represented across items.
√	The item is free of content that might offend a geographic subgroup.
√	The item is free of content that might unfairly advantage or disadvantage a geographic subgroup.
G.	Disability Bias
Definition	All groups of society should be portrayed accurately and fairly regarding disability. Stereotypes related to any particular disability should be avoided. No undue restrictions should exist in the item that would interfere with the ability of a student to comprehend or respond to the item.
√	The item is free of content that might offend a disability subgroup.
√	The item is free of content that might unfairly advantage or disadvantage a disability subgroup.
√	A graphic representation is used in the items, as appropriate. The complexity of the graphic is appropriate to the purpose—simpler is better.
√	The item avoids content that depends on sensory knowledge (such as references to movement, sound, smell, etc.) unless this is crucial to the overall item.
√	The item could be put into Braille.
√	The item avoids using both O and Q.
√	Letter pairs can be easily distinguished when read. (S and T are okay; S and X are not).
H.	Art Supports Text
Definition	The art is related to the item and supports the reader when possible. The item text and art are legible and accessible, and the art is appropriately placed in the item to support the reader. The art does not distract the test taker, but instead provides a scaffold to overall comprehension.
√	All pictures relate to items.
√	The item is free from pictorial clutter: All pictures are needed to answer the item.
√	Graphics are clear and non-fuzzy.
√	Any symbols used are highly distinguishable.
√	Visual load requirements are reasonable for the grade level.
√	Multi-dimensional graphics and complex shading are avoided.
√	Tables have replaced any cluttered graphs.
√	Labels read clockwise (as is easier for Braille readers).
I.	Special Populations Considerations
Definition	Consideration must be given for maximum accessibility to all students including but not limited to English language learners, limited sight, hearing impaired, cognitively challenged, etc. These considerations will assist all students.
√	The item contains scaffolding techniques to support student understanding of what is being asked in the item.
√	Text is replaced with graphic representations, when appropriate.
√	The item is written with simplified text load.

Universal Design Item Checklist	
√	The item is written with simplified sentences.
√	The item has as little extraneous information as possible.
√	The item provides context, but it is simplified.
√	The item uses smaller or less complicated numbers or expressions where not otherwise required.
√	The item avoids negative phrasing or questions; for example, questions are not asked in the negative.

Appendix E: Criteria for Item Acceptability

The following criteria represent best practices in item development, and were implemented during the creation and review of the NYS 3–8 CCSS test questions; however, these criteria are not a substitute for the full, detailed criteria documents which are available online at the following links:

- <http://www.engageny.org/resource/new-york-state-item-review-criteria-for-grade-3-8-english-language-arts-tests>; and
- <http://www.engageny.org/resource/new-york-state-item-review-criteria-for-grade-3-8-mathematics-tests>.

For Multiple-Choice Items:

Check that the content of each item

- is targeted to assess only one objective or skill (unless specifications indicate otherwise)
- deals with material that is important in testing the targeted performance indicator
- uses grade-appropriate content and thinking skills
- is presented at a reading level suitable for the grade level being tested
- has a stem that facilitates answering the question or completing the statement without looking at the answer choices
- has a stem that does **not** present clues to the correct answer choice
- has answer choices that are plausible and attractive to the student who has not mastered the objective or skill
- has mutually exclusive distractors
- has one and only one correct answer choice
- is free of cultural, racial, ethnic, age, gender, disability, regional, or other apparent bias

Check that the format of each item

- is worded in the positive unless it is absolutely necessary to use the negative form
- is free of extraneous words or expressions in both the stem and the answer choices (e.g., the same word or phrase does not begin each answer choice)
- indicates emphasis on key words, such as best, first, least, not, and others that are important and might be overlooked
- places the interrogative word at the **beginning** of a stem in the form of a question, or places the omitted portion of an incomplete statement at the **end** of the statement
- indicates the correct answer choice
- provides the rationale for all distractors
- is conceptually, grammatically, and syntactically consistent—between the stem and answer choices, and among the answer choices
- has answer choices balanced in length, or contains two long and two short answer choices
- clearly identifies the passage or other stimulus material associated with the item
- clearly identifies a need of art, if applicable, and the art is conceptualized and sketched, with important considerations explicated

Also check that

- one item does not present clues to the correct answer choice for any other item
- any item based on a passage is answerable from the information given in the passage and is not dependent on skills related to other content areas
- any item based on a passage is truly passage-dependent; that is, **not** answerable without reference to the passage
- there is a balance of reasonable, non-stereotypical representation of economic classes, races, cultures, ages, genders, and persons with disabilities in context and art

For Constructed-Response Items:**Check that the content of each item is**

- designed to assess the targeted performance indicator
- appropriate for the grade level being tested
- presented at a reading level suitable for the grade level being tested
- appropriate in context
- written so that a student possessing knowledge or skill being tested can construct a response that can be scored with the specified rubric or scoring tool; that is, the range of possible correct responses must be wide enough to allow for a diversity of responses, but narrow enough so that students who do not clearly show their grasp of the objective or skill being assessed cannot obtain the maximum score
- presented without clues to the correct response
- checked for accuracy and documented against reliable, up-to-date sources (including rubrics)
- free of cultural, racial, ethnic, age, gender, disability, or other apparent bias

Check that the format of each item is

- appropriate for the question being asked and the intended response
- worded clearly and concisely, using simple vocabulary and sentence structure
- precise and unambiguous in its directions for the desired response
- free of extraneous words or expressions
- worded in the positive form rather than in the negative form
- conceptually, grammatically, and syntactically consistent
- marked with emphasis on key words, such as best, first, least, and others that are important and might be overlooked
- clearly identified as needing art, if applicable, and the art is conceptualized and sketched, with important considerations explicated

Also check that

- one item does not present clues to the correct response to any other item
- there is a balance of reasonable, non-stereotypical representation of economic classes, races, cultures, ages, genders, and persons with disabilities in context and art
- for each set of items related to a reading passage, each item is designed to elicit a unique and independent response
- items designed to assess reading do not depend on prior knowledge of the subject matter used in the prompt/question

Appendix F: Psychometric Guidelines for Operational Item Selection

It is primarily up to the content development department to select items for the 2014 Common Core Operational Test. The psychometrics department will provide support, as necessary, and will review the final item selection. The psychometrics department will provide data files with parameters for all FT items eligible for the item pool. The pools of items eligible for 2014 item selection included 2013 embedded and stand-alone field test items and items field-tested in New York State in 2013.

Here are the general guidelines for item selection:

- Satisfy the content specifications in terms of objective coverage and the number and percentage of MC and CR items on the test. An often-used criterion for objective coverage is within 5% of the percentages of score points and items per objective.
- To the extent possible, select both easy and difficult items to provide good measurement information at both ends of the performance scale.
- Avoid selecting items with too high/low p-values, items with flagged point biserials, and poorly fitting items.
- Minimize the number of items flagged for DIF (gender, ethnic, and High/Low Needs schools). Flagged items should be reviewed for content again. It needs to be remembered that some items may be flagged for DIF by chance only, and their content may not necessarily be biased against any of the analyzed subgroups. The psychometrics department will provide DIF information for each item. It is also possible to get “significant” DIF, but not bias, if the content is a necessary part of the construct that is measured. That is, some items may be flagged for DIF not out of chance and still not represent bias.
- Provide the NYSED with the following summary information:
 - Overview of the statistical properties of the tests
 - Blueprint comparison between the test build and the target. The focus is on the total number of points on the test
 - Raw score proportion correct comparison between the test build and the reference (i.e., Spring 2013 test)
 - Vertical linked average difficulty parameter (MC items only) across all grades
 - Vertically linked TCC based on the constructed test
 - TCC, Test Information Curves and Conditional SEM Curves for each subject and grade, again using Spring 2013 operational test as a reference.

Appendix G: Operational Item Maps

The following tables show the operational item maps for the 2014 NYSTP Grades 3–8 Common Core ELA and Mathematics Tests. External linking and field-test items (i.e., those not contributing to students’ scores) have been omitted. Additional detail on the standards to which these items align may be found at: <http://www.engageny.org/resource/new-york-state-p-12-common-core-learning-standards>.

Table G1. ELA Grade 3 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RL.3.4
2	MC	1	CCSS.ELA-Literacy.RL.3.1
3	MC	1	CCSS.ELA-Literacy.RL.3.3
4	MC	1	CCSS.ELA-Literacy.RL.3.1
5	MC	1	CCSS.ELA-Literacy.RL.3.1
6	MC	1	CCSS.ELA-Literacy.RL.3.2
7	MC	1	CCSS.ELA-Literacy.RI.3.1
8	MC	1	CCSS.ELA-Literacy.RI.3.8
9	MC	1	CCSS.ELA-Literacy.RI.3.1
10	MC	1	CCSS.ELA-Literacy.RI.3.7
11	MC	1	CCSS.ELA-Literacy.RI.3.4
12	MC	1	CCSS.ELA-Literacy.RI.3.3
13	MC	1	CCSS.ELA-Literacy.RL.3.1
14	MC	1	CCSS.ELA-Literacy.RL.3.1
15	MC	1	CCSS.ELA-Literacy.RL.3.1
16	MC	1	CCSS.ELA-Literacy.RL.3.3
17	MC	1	CCSS.ELA-Literacy.RL.3.5
18	MC	1	CCSS.ELA-Literacy.RL.3.2
19	MC	1	CCSS.ELA-Literacy.RI.3.5
20	MC	1	CCSS.ELA-Literacy.RI.3.3
21	MC	1	CCSS.ELA-Literacy.RI.3.1
22	MC	1	CCSS.ELA-Literacy.RI.3.5
23	MC	1	CCSS.ELA-Literacy.RI.3.4
24	MC	1	CCSS.ELA-Literacy.RI.3.1
25	MC	1	CCSS.ELA-Literacy.L.3.4
26	MC	1	CCSS.ELA-Literacy.RI.3.8
27	MC	1	CCSS.ELA-Literacy.RI.3.2
28	MC	1	CCSS.ELA-Literacy.RI.3.5
29	MC	1	CCSS.ELA-Literacy.RI.3.7
30	CR	2	CCSS.ELA-Literacy.RL.3.3
31	CR	2	CCSS.ELA-Literacy.RL.3.1
32	CR	2	CCSS.ELA-Literacy.RI.3.5
33	CR	4	CCSS.ELA-Literacy.RI.3.3

Table G1. ELA Grade 3 Operational Item Map (cont.)

Item	Type	Points	Standard
34	CR	2	CCSS.ELA-Literacy.RI.3.3
35	CR	2	CCSS.ELA-Literacy.RI.3.8
36	CR	2	CCSS.ELA-Literacy.RL.3.3
37	CR	2	CCSS.ELA-Literacy.RL.3.1
38	CR	2	CCSS.ELA-Literacy.RL.3.1

Table G2. ELA Grade 4 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RL.4.1
2	MC	1	CCSS.ELA-Literacy.RL.4.1
3	MC	1	CCSS.ELA-Literacy.RL.4.2
4	MC	1	CCSS.ELA-Literacy.RL.4.3
5	MC	1	CCSS.ELA-Literacy.RL.4.2
6	MC	1	CCSS.ELA-Literacy.RL.4.6
7	MC	1	CCSS.ELA-Literacy.RI.4.1
8	MC	1	CCSS.ELA-Literacy.RI.4.4
9	MC	1	CCSS.ELA-Literacy.RI.4.1
10	MC	1	CCSS.ELA-Literacy.RI.4.8
11	MC	1	CCSS.ELA-Literacy.RI.4.1
12	MC	1	CCSS.ELA-Literacy.RI.4.2
13	MC	1	CCSS.ELA-Literacy.RL.4.4
14	MC	1	CCSS.ELA-Literacy.RL.4.1
15	MC	1	CCSS.ELA-Literacy.RL.4.3
16	MC	1	CCSS.ELA-Literacy.RL.4.4
17	MC	1	CCSS.ELA-Literacy.RL.4.7
18	MC	1	CCSS.ELA-Literacy.RL.4.3
19	MC	1	CCSS.ELA-Literacy.RL.4.1
20	MC	1	CCSS.ELA-Literacy.RL.4.4
21	MC	1	CCSS.ELA-Literacy.RL.4.1
22	MC	1	CCSS.ELA-Literacy.L.4.5a
23	MC	1	CCSS.ELA-Literacy.RL.4.2
24	MC	1	CCSS.ELA-Literacy.RL.4.5
25	MC	1	CCSS.ELA-Literacy.RI.4.5
26	MC	1	CCSS.ELA-Literacy.RI.4.8
27	MC	1	CCSS.ELA-Literacy.RI.4.2
28	MC	1	CCSS.ELA-Literacy.RI.4.3
29	MC	1	CCSS.ELA-Literacy.RI.4.3
30	MC	1	CCSS.ELA-Literacy.L.4.4a
31	MC	1	CCSS.ELA-Literacy.RI.4.1

Table G2. ELA Grade 4 Operational Item Map (cont.)

Item	Type	Points	Standard
32	CR	2	CCSS.ELA-Literacy.RI.4.2
33	CR	2	CCSS.ELA-Literacy.RI.4.8
34	CR	2	CCSS.ELA-Literacy.RL.4.3
35	CR	4	CCSS.ELA-Literacy.RL.4.2
36	CR	2	CCSS.ELA-Literacy.RL.4.3
37	CR	2	CCSS.ELA-Literacy.RL.4.2
38	CR	2	CCSS.ELA-Literacy.RI.4.8
39	CR	2	CCSS.ELA-Literacy.RI.4.5
40	CR	2	CCSS.ELA-Literacy.RI.4.3
41	CR	4	CCSS.ELA-Literacy.RI.4.9

Table G3. ELA Grade 5 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.L.5.4a
2	MC	1	CCSS.ELA-Literacy.RL.5.3
3	MC	1	CCSS.ELA-Literacy.RL.5.4
4	MC	1	CCSS.ELA-Literacy.RL.5.3
5	MC	1	CCSS.ELA-Literacy.RL.5.4
6	MC	1	CCSS.ELA-Literacy.RL.5.1
7	MC	1	CCSS.ELA-Literacy.RL.5.2
8	MC	1	CCSS.ELA-Literacy.RI.5.5
9	MC	1	CCSS.ELA-Literacy.RI.5.2
10	MC	1	CCSS.ELA-Literacy.RI.5.4
11	MC	1	CCSS.ELA-Literacy.RI.5.8
12	MC	1	CCSS.ELA-Literacy.RI.5.3
13	MC	1	CCSS.ELA-Literacy.RI.5.1
14	MC	1	CCSS.ELA-Literacy.RI.5.1
15	MC	1	CCSS.ELA-Literacy.RL.5.1
16	MC	1	CCSS.ELA-Literacy.RL.5.3
17	MC	1	CCSS.ELA-Literacy.RL.5.6
18	MC	1	CCSS.ELA-Literacy.RL.5.2
19	MC	1	CCSS.ELA-Literacy.RL.5.5
20	MC	1	CCSS.ELA-Literacy.RL.5.3
21	MC	1	CCSS.ELA-Literacy.RL.5.1
22	MC	1	CCSS.ELA-Literacy.RI.5.4
23	MC	1	CCSS.ELA-Literacy.RI.5.2
24	MC	1	CCSS.ELA-Literacy.RI.5.2
25	MC	1	CCSS.ELA-Literacy.RI.5.8
26	MC	1	CCSS.ELA-Literacy.RI.5.1
27	MC	1	CCSS.ELA-Literacy.RI.5.3

Table G3. ELA Grade 5 Operational Item Map (cont.)

Item	Type	Points	Standard
28	MC	1	CCSS.ELA-Literacy.RI.5.3
29	MC	1	CCSS.ELA-Literacy.RL.5.4
30	MC	1	CCSS.ELA-Literacy.RL.5.2
31	MC	1	CCSS.ELA-Literacy.RL.5.5
32	MC	1	CCSS.ELA-Literacy.RL.5.4
33	MC	1	CCSS.ELA-Literacy.RL.5.2
34	MC	1	CCSS.ELA-Literacy.RL.5.5
35	MC	1	CCSS.ELA-Literacy.RL.5.1
36	MC	1	CCSS.ELA-Literacy.RI.5.4
37	MC	1	CCSS.ELA-Literacy.RI.5.7
38	MC	1	CCSS.ELA-Literacy.RI.5.8
39	MC	1	CCSS.ELA-Literacy.RI.5.4
40	MC	1	CCSS.ELA-Literacy.RI.5.1
41	MC	1	CCSS.ELA-Literacy.RI.5.3
42	MC	1	CCSS.ELA-Literacy.RI.5.3
43	CR	2	CCSS.ELA-Literacy.RI.5.1
44	CR	2	CCSS.ELA-Literacy.RI.5.2
45	CR	2	CCSS.ELA-Literacy.RL.5.2
46	CR	4	CCSS.ELA-Literacy.RL.5.3
47	CR	2	CCSS.ELA-Literacy.RL.5.5
48	CR	2	CCSS.ELA-Literacy.RL.5.6
49	CR	2	CCSS.ELA-Literacy.RI.5.2
50	CR	2	CCSS.ELA-Literacy.RI.5.3
51	CR	2	CCSS.ELA-Literacy.RI.5.2
52	CR	4	CCSS.ELA-Literacy.RI.5.9

Table G4. ELA Grade 6 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RI.6.1
2	MC	1	CCSS.ELA-Literacy.RI.6.1
3	MC	1	CCSS.ELA-Literacy.RI.6.4
4	MC	1	CCSS.ELA-Literacy.RI.6.5
5	MC	1	CCSS.ELA-Literacy.RI.6.3
6	MC	1	CCSS.ELA-Literacy.RI.6.8
7	MC	1	CCSS.ELA-Literacy.RI.6.6
8	MC	1	CCSS.ELA-Literacy.RI.6.5
9	MC	1	CCSS.ELA-Literacy.RI.6.5
10	MC	1	CCSS.ELA-Literacy.RI.6.4
11	MC	1	CCSS.ELA-Literacy.RI.6.3
12	MC	1	CCSS.ELA-Literacy.RI.6.2

Table G4. ELA Grade 6 Operational Item Map (cont.)

Item	Type	Points	Standard
13	MC	1	CCSS.ELA-Literacy.RI.6.1
14	MC	1	CCSS.ELA-Literacy.RI.6.2
15	MC	1	CCSS.ELA-Literacy.RL.6.4
16	MC	1	CCSS.ELA-Literacy.RL.6.2
17	MC	1	CCSS.ELA-Literacy.RL.6.1
18	MC	1	CCSS.ELA-Literacy.RL.6.5
19	MC	1	CCSS.ELA-Literacy.RL.6.4
20	MC	1	CCSS.ELA-Literacy.RL.6.2
21	MC	1	CCSS.ELA-Literacy.RI.6.2
22	MC	1	CCSS.ELA-Literacy.RI.6.5
23	MC	1	CCSS.ELA-Literacy.RI.6.1
24	MC	1	CCSS.ELA-Literacy.RI.6.3
25	MC	1	CCSS.ELA-Literacy.RI.6.1
26	MC	1	CCSS.ELA-Literacy.RI.6.6
27	MC	1	CCSS.ELA-Literacy.RI.6.2
28	MC	1	CCSS.ELA-Literacy.RI.6.7
29	MC	1	CCSS.ELA-Literacy.RI.6.4
30	MC	1	CCSS.ELA-Literacy.RI.6.1
31	MC	1	CCSS.ELA-Literacy.RI.6.5
32	MC	1	CCSS.ELA-Literacy.RI.6.1
33	MC	1	CCSS.ELA-Literacy.L.6.4a
34	MC	1	CCSS.ELA-Literacy.RI.6.5
35	MC	1	CCSS.ELA-Literacy.RI.6.2
36	MC	1	CCSS.ELA-Literacy.RL.6.3
37	MC	1	CCSS.ELA-Literacy.RL.6.5
38	MC	1	CCSS.ELA-Literacy.RL.6.3
39	MC	1	CCSS.ELA-Literacy.RL.6.4
40	MC	1	CCSS.ELA-Literacy.RL.6.6
41	MC	1	CCSS.ELA-Literacy.RL.6.1
42	MC	1	CCSS.ELA-Literacy.L.6.4a
43	CR	2	CCSS.ELA-Literacy.RI.6.4
44	CR	2	CCSS.ELA-Literacy.RI.6.1
45	CR	2	CCSS.ELA-Literacy.RI.6.1
46	CR	4	CCSS.ELA-Literacy.RI.6.9
47	CR	2	CCSS.ELA-Literacy.RI.6.3
48	CR	2	CCSS.ELA-Literacy.RI.6.4
49	CR	2	CCSS.ELA-Literacy.RL.6.2
50	CR	2	CCSS.ELA-Literacy.RL.6.1
51	CR	2	CCSS.ELA-Literacy.RL.6.5
52	CR	4	CCSS.ELA-Literacy.RL.6.2

Table G5. ELA Grade 7 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RI.7.6
2	MC	1	CCSS.ELA-Literacy.RI.7.1
3	MC	1	CCSS.ELA-Literacy.RI.7.5
4	MC	1	CCSS.ELA-Literacy.RI.7.1
5	MC	1	CCSS.ELA-Literacy.RI.7.4
6	MC	1	CCSS.ELA-Literacy.L.7.4a
7	MC	1	CCSS.ELA-Literacy.RI.7.3
8	MC	1	CCSS.ELA-Literacy.RL.7.3
9	MC	1	CCSS.ELA-Literacy.RL.7.2
10	MC	1	CCSS.ELA-Literacy.RL.7.1
11	MC	1	CCSS.ELA-Literacy.RL.7.1
12	MC	1	CCSS.ELA-Literacy.RL.7.5
13	MC	1	CCSS.ELA-Literacy.RI.7.1
14	MC	1	CCSS.ELA-Literacy.RI.7.3
15	MC	1	CCSS.ELA-Literacy.RI.7.4
16	MC	1	CCSS.ELA-Literacy.RI.7.8
17	MC	1	CCSS.ELA-Literacy.RI.7.2
18	MC	1	CCSS.ELA-Literacy.RI.7.8
19	MC	1	CCSS.ELA-Literacy.RI.7.5
20	MC	1	CCSS.ELA-Literacy.RL.7.3
21	MC	1	CCSS.ELA-Literacy.RL.7.6
22	MC	1	CCSS.ELA-Literacy.RL.7.3
23	MC	1	CCSS.ELA-Literacy.RL.7.4
24	MC	1	CCSS.ELA-Literacy.RL.7.1
25	MC	1	CCSS.ELA-Literacy.RL.7.4
26	MC	1	CCSS.ELA-Literacy.RL.7.3
27	MC	1	CCSS.ELA-Literacy.RL.7.6
28	MC	1	CCSS.ELA-Literacy.RI.7.3
29	MC	1	CCSS.ELA-Literacy.RI.7.4
30	MC	1	CCSS.ELA-Literacy.RI.7.3
31	MC	1	CCSS.ELA-Literacy.RI.7.1
32	MC	1	CCSS.ELA-Literacy.RI.7.1
33	MC	1	CCSS.ELA-Literacy.RI.7.5
34	MC	1	CCSS.ELA-Literacy.RI.7.2
35	MC	1	CCSS.ELA-Literacy.RI.7.4
36	MC	1	CCSS.ELA-Literacy.RI.7.5
37	MC	1	CCSS.ELA-Literacy.RI.7.1
38	MC	1	CCSS.ELA-Literacy.RI.7.1
39	MC	1	CCSS.ELA-Literacy.RI.7.6
40	MC	1	CCSS.ELA-Literacy.RI.7.8
41	MC	1	CCSS.ELA-Literacy.RI.7.7

Table G5. ELA Grade 7 Operational Item Map (cont.)

Item	Type	Points	Standard
42	CR	2	CCSS.ELA-Literacy.RI.7.4
43	CR	2	CCSS.ELA-Literacy.RI.7.6
44	CR	2	CCSS.ELA-Literacy.RL.7.5
45	CR	4	CCSS.ELA-Literacy.RL.7.2
46	CR	2	CCSS.ELA-Literacy.RI.7.5
47	CR	2	CCSS.ELA-Literacy.RI.7.8
48	CR	2	CCSS.ELA-Literacy.RL.7.5
49	CR	2	CCSS.ELA-Literacy.RI.7.5
50	CR	2	CCSS.ELA-Literacy.RI.7.2
51	CR	4	CCSS.ELA-Literacy.RL.7.9

Table G6. ELA Grade 8 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RI.8.4
2	MC	1	CCSS.ELA-Literacy.RI.8.5
3	MC	1	CCSS.ELA-Literacy.RI.8.3
4	MC	1	CCSS.ELA-Literacy.RI.8.2
5	MC	1	CCSS.ELA-Literacy.RI.8.8
6	MC	1	CCSS.ELA-Literacy.RI.8.5
7	MC	1	CCSS.ELA-Literacy.RI.8.6
8	MC	1	CCSS.ELA-Literacy.RL.8.1
9	MC	1	CCSS.ELA-Literacy.RL.8.3
10	MC	1	CCSS.ELA-Literacy.RL.8.6
11	MC	1	CCSS.ELA-Literacy.RL.8.4
12	MC	1	CCSS.ELA-Literacy.RL.8.2
13	MC	1	CCSS.ELA-Literacy.RL.8.3
14	MC	1	CCSS.ELA-Literacy.RL.8.1
15	MC	1	CCSS.ELA-Literacy.RI.8.5
16	MC	1	CCSS.ELA-Literacy.RI.8.4
17	MC	1	CCSS.ELA-Literacy.RI.8.2
18	MC	1	CCSS.ELA-Literacy.RI.8.2
19	MC	1	CCSS.ELA-Literacy.L.8.4a
20	MC	1	CCSS.ELA-Literacy.RI.8.8
21	MC	1	CCSS.ELA-Literacy.RI.8.2
22	MC	1	CCSS.ELA-Literacy.RL.8.1
23	MC	1	CCSS.ELA-Literacy.RL.8.4
24	MC	1	CCSS.ELA-Literacy.RL.8.3
25	MC	1	CCSS.ELA-Literacy.RL.8.1
26	MC	1	CCSS.ELA-Literacy.RL.8.6
27	MC	1	CCSS.ELA-Literacy.RL.8.3

Table G6. ELA Grade 8 Operational Item Map (cont.)

Item	Type	Points	Standard
28	MC	1	CCSS.ELA-Literacy.L.8.4a
29	MC	1	CCSS.ELA-Literacy.RI.8.5
30	MC	1	CCSS.ELA-Literacy.RI.8.4
31	MC	1	CCSS.ELA-Literacy.RI.8.3
32	MC	1	CCSS.ELA-Literacy.RI.8.2
33	MC	1	CCSS.ELA-Literacy.RI.8.6
34	MC	1	CCSS.ELA-Literacy.RI.8.1
35	MC	1	CCSS.ELA-Literacy.RI.8.2
36	MC	1	CCSS.ELA-Literacy.RL.8.4
37	MC	1	CCSS.ELA-Literacy.RL.8.3
38	MC	1	CCSS.ELA-Literacy.RL.8.4
39	MC	1	CCSS.ELA-Literacy.RL.8.2
40	MC	1	CCSS.ELA-Literacy.RL.8.2
41	MC	1	CCSS.ELA-Literacy.RL.8.2
42	MC	1	CCSS.ELA-Literacy.RL.8.6
43	CR	2	CCSS.ELA-Literacy.RI.8.5
44	CR	2	CCSS.ELA-Literacy.RI.8.3
45	CR	2	CCSS.ELA-Literacy.RI.8.3
46	CR	4	CCSS.ELA-Literacy.RI.8.8
47	CR	2	CCSS.ELA-Literacy.RI.8.1
48	CR	2	CCSS.ELA-Literacy.RI.8.6
49	CR	2	CCSS.ELA-Literacy.RI.8.1
50	CR	2	CCSS.ELA-Literacy.RL.8.6
51	CR	2	CCSS.ELA-Literacy.RL.8.4
52	CR	4	CCSS.ELA-Literacy.RL.8.2

Table G7. Mathematics Grade 3 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.3.OA.A.1
2	MC	1	CCSS.Math.Content.3.MD.C.6
3	MC	1	CCSS.Math.Content.3.NF.A.1
4	MC	1	CCSS.Math.Content.3.OA.B.6
5	MC	1	CCSS.Math.Content.3.G.A.2
6	MC	1	CCSS.Math.Content.3.OA.A.3
7	MC	1	CCSS.Math.Content.3.NF.A.3b
8	MC	1	CCSS.Math.Content.3.OA.A.3
9	MC	1	CCSS.Math.Content.3.MD.A.1
10	MC	1	CCSS.Math.Content.3.OA.A.4
11	MC	1	CCSS.Math.Content.3.MD.C.7b
12	MC	1	CCSS.Math.Content.3.NF.A.2b

Table G7. Mathematics Grade 3 Operational Item Map (cont.)

Item	Type	Points	Standard
13	MC	1	CCSS.Math.Content.3.MD.C.7d
14	MC	1	CCSS.Math.Content.3.OA.D.8
15	MC	1	CCSS.Math.Content.3.MD.B.3
16	MC	1	CCSS.Math.Content.3.OA.D.9
17	MC	1	CCSS.Math.Content.3.OA.B.5
18	MC	1	CCSS.Math.Content.3.MD.C.7c
19	MC	1	CCSS.Math.Content.3.OA.D.8
20	MC	1	CCSS.Math.Content.3.MD.C.5a
21	MC	1	CCSS.Math.Content.3.OA.B.6
22	MC	1	CCSS.Math.Content.3.NF.A.1
23	MC	1	CCSS.Math.Content.3.OA.A.3
24	MC	1	CCSS.Math.Content.3.NF.A.2b
25	MC	1	CCSS.Math.Content.3.MD.C.7b
26	MC	1	CCSS.Math.Content.3.NF.A.1
27	MC	1	CCSS.Math.Content.3.MD.A.2
28	MC	1	CCSS.Math.Content.3.MD.A.1
29	MC	1	CCSS.Math.Content.3.MD.A.2
30	MC	1	CCSS.Math.Content.3.MD.C.7a
31	MC	1	CCSS.Math.Content.3.OA.B.5
32	MC	1	CCSS.Math.Content.3.NBT.A.1
33	MC	1	CCSS.Math.Content.3.NF.A.2b
34	MC	1	CCSS.Math.Content.3.OA.A.4
35	MC	1	CCSS.Math.Content.3.NF.A.3a
36	MC	1	CCSS.Math.Content.3.G.A.2
37	MC	1	CCSS.Math.Content.3.OA.D.9
38	MC	1	CCSS.Math.Content.3.NF.A.3a
39	MC	1	CCSS.Math.Content.3.OA.A.1
40	MC	1	CCSS.Math.Content.3.MD.C.6
41	MC	1	CCSS.Math.Content.3.NF.A.3b
42	CR	2	CCSS.Math.Content.3.OA.A.2
43	CR	2	CCSS.Math.Content.3.OA.D.8
44	CR	2	CCSS.Math.Content.3.NF.A.3b
45	CR	2	CCSS.Math.Content.3.OA.A.3
46	CR	2	CCSS.Math.Content.3.MD.B.3
47	CR	3	CCSS.Math.Content.3.OA.A.3
48	CR	3	CCSS.Math.Content.3.OA.D.8
49	CR	3	CCSS.Math.Content.3.NBT.A.3

Table G8. Mathematics Grade 4 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.4.NF.A.1
2	MC	1	CCSS.Math.Content.3.MD.B.4
3	MC	1	CCSS.Math.Content.4.NBT.A.3
4	MC	1	CCSS.Math.Content.3.MD.D.8
5	MC	1	CCSS.Math.Content.4.G.A.1
6	MC	1	CCSS.Math.Content.4.NF.B.3a
7	MC	1	CCSS.Math.Content.4.OA.B.4
8	MC	1	CCSS.Math.Content.4.NBT.A.1
9	MC	1	CCSS.Math.Content.4.NF.B.4a
10	MC	1	CCSS.Math.Content.4.OA.C.5
11	MC	1	CCSS.Math.Content.4.NF.A.2
12	MC	1	CCSS.Math.Content.4.NBT.B.5
13	MC	1	CCSS.Math.Content.4.NBT.A.2
14	MC	1	CCSS.Math.Content.4.MD.C.7
15	MC	1	CCSS.Math.Content.4.NF.A.1
16	MC	1	CCSS.Math.Content.4.NBT.B.6
17	MC	1	CCSS.Math.Content.4.G.A.3
18	MC	1	CCSS.Math.Content.4.MD.C.6
19	MC	1	CCSS.Math.Content.4.NBT.B.5
20	MC	1	CCSS.Math.Content.4.MD.C.5a
21	MC	1	CCSS.Math.Content.4.G.A.2
22	MC	1	CCSS.Math.Content.4.NBT.A.2
23	MC	1	CCSS.Math.Content.4.NF.A.1
24	MC	1	CCSS.Math.Content.4.MD.C.6
25	MC	1	CCSS.Math.Content.4.NBT.A.2
26	MC	1	CCSS.Math.Content.4.NF.B.3b
27	MC	1	CCSS.Math.Content.4.NBT.B.5
28	MC	1	CCSS.Math.Content.4.G.A.2
29	MC	1	CCSS.Math.Content.4.MD.C.7
30	MC	1	CCSS.Math.Content.4.NF.B.3c
31	MC	1	CCSS.Math.Content.4.NBT.A.1
32	MC	1	CCSS.Math.Content.4.OA.A.2
33	MC	1	CCSS.Math.Content.4.NF.A.1
34	MC	1	CCSS.Math.Content.4.NF.B.4c
35	MC	1	CCSS.Math.Content.4.NBT.B.6
36	MC	1	CCSS.Math.Content.4.NF.A.2
37	MC	1	CCSS.Math.Content.4.OA.A.1
38	MC	1	CCSS.Math.Content.4.OA.A.3
39	MC	1	CCSS.Math.Content.4.MD.A.3
40	MC	1	CCSS.Math.Content.4.NBT.A.1
41	MC	1	CCSS.Math.Content.4.MD.C.5b

Table G8. Mathematics Grade 4 Operational Item Map (cont.)

Item	Type	Points	Standard
42	MC	1	CCSS.Math.Content.4.MD.B.4
43	CR	2	CCSS.Math.Content.4.NF.A.2
44	CR	2	CCSS.Math.Content.4.OA.A.2
45	CR	2	CCSS.Math.Content.4.NF.B.3d
46	CR	2	CCSS.Math.Content.4.G.A.1
47	CR	2	CCSS.Math.Content.4.MD.A.3
48	CR	2	CCSS.Math.Content.4.NBT.B.6
49	CR	3	CCSS.Math.Content.4.OA.A.2
50	CR	3	CCSS.Math.Content.4.OA.A.3
51	CR	3	CCSS.Math.Content.4.NF.B.4c
52	CR	3	CCSS.Math.Content.4.NBT.B.5

Table G9. Mathematics Grade 5 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.5.OA.A.2
2	MC	1	CCSS.Math.Content.5.MD.C.3a
3	MC	1	CCSS.Math.Content.5.NBT.B.6
4	MC	1	CCSS.Math.Content.4.NF.C.7
5	MC	1	CCSS.Math.Content.5.MD.C.3b
6	MC	1	CCSS.Math.Content.5.MD.C.5b
7	MC	1	CCSS.Math.Content.5.NF.B.6
8	MC	1	CCSS.Math.Content.5.NBT.B.6
9	MC	1	CCSS.Math.Content.5.NF.B.7c
10	MC	1	CCSS.Math.Content.5.NBT.A.1
11	MC	1	CCSS.Math.Content.5.OA.A.1
12	MC	1	CCSS.Math.Content.5.NF.B.6
13	MC	1	CCSS.Math.Content.5.MD.A.1
14	MC	1	CCSS.Math.Content.5.NF.B.4b
15	MC	1	CCSS.Math.Content.5.NF.B.5a
16	MC	1	CCSS.Math.Content.5.NBT.B.6
17	MC	1	CCSS.Math.Content.5.MD.C.4
18	MC	1	CCSS.Math.Content.5.NF.B.7c
19	MC	1	CCSS.Math.Content.5.NF.B.4b
20	MC	1	CCSS.Math.Content.5.NF.A.1
21	MC	1	CCSS.Math.Content.5.G.B.3
22	MC	1	CCSS.Math.Content.5.NF.B.5a
23	MC	1	CCSS.Math.Content.5.OA.A.2
24	MC	1	CCSS.Math.Content.4.MD.A.2
25	MC	1	CCSS.Math.Content.5.NBT.A.4
26	MC	1	CCSS.Math.Content.5.MD.C.3b

Table G9. Mathematics Grade 5 Operational Item Map (cont.)

Item	Type	Points	Standard
27	MC	1	CCSS.Math.Content.5.G.B.4
28	MC	1	CCSS.Math.Content.5.NBT.A.2
29	MC	1	CCSS.Math.Content.5.MD.A.1
30	MC	1	CCSS.Math.Content.5.NF.B.3
31	MC	1	CCSS.Math.Content.5.NF.B.7a
32	MC	1	CCSS.Math.Content.5.MD.A.1
33	MC	1	CCSS.Math.Content.4.NF.C.6
34	MC	1	CCSS.Math.Content.5.NF.B.6
35	MC	1	CCSS.Math.Content.5.NF.B.7c
36	MC	1	CCSS.Math.Content.5.NBT.A.1
37	MC	1	CCSS.Math.Content.5.MD.A.1
38	MC	1	CCSS.Math.Content.5.NF.A.2
39	MC	1	CCSS.Math.Content.5.NF.B.3
40	MC	1	CCSS.Math.Content.5.NF.B.5a
41	MC	1	CCSS.Math.Content.5.OA.A.1
42	MC	1	CCSS.Math.Content.5.NF.A.1
43	CR	2	CCSS.Math.Content.5.NBT.B.7
44	CR	2	CCSS.Math.Content.5.NBT.A.3b
45	CR	2	CCSS.Math.Content.5.MD.B.2
46	CR	2	CCSS.Math.Content.5.NBT.B.6
47	CR	2	CCSS.Math.Content.5.MD.C.4
48	CR	2	CCSS.Math.Content.5.NF.A.2
49	CR	3	CCSS.Math.Content.5.MD.C.5a
50	CR	3	CCSS.Math.Content.5.NBT.B.7
51	CR	3	CCSS.Math.Content.5.NF.B.6
52	CR	3	CCSS.Math.Content.5.NF.A.2

Table G10. Mathematics Grade 6 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.6.EE.A.1
2	MC	1	CCSS.Math.Content.6.EE.B.7
3	MC	1	CCSS.Math.Content.6.RP.A.1
4	MC	1	CCSS.Math.Content.6.NS.C.7d
5	MC	1	CCSS.Math.Content.6.EE.B.5
6	MC	1	CCSS.Math.Content.6.NS.A.1
7	MC	1	CCSS.Math.Content.6.EE.B.8
8	MC	1	CCSS.Math.Content.6.EE.A.1
9	MC	1	CCSS.Math.Content.6.EE.A.2c
10	MC	1	CCSS.Math.Content.6.EE.A.3
11	MC	1	CCSS.Math.Content.6.RP.A.2
12	MC	1	CCSS.Math.Content.6.G.A.4
13	MC	1	CCSS.Math.Content.6.NS.A.1
14	MC	1	CCSS.Math.Content.6.EE.B.7
15	MC	1	CCSS.Math.Content.6.EE.A.3
16	MC	1	CCSS.Math.Content.5.OA.B.3
17	MC	1	CCSS.Math.Content.6.NS.B.4
18	MC	1	CCSS.Math.Content.6.G.A.1
19	MC	1	CCSS.Math.Content.6.EE.B.7
20	MC	1	CCSS.Math.Content.6.RP.A.3b
21	MC	1	CCSS.Math.Content.6.NS.A.1
22	MC	1	CCSS.Math.Content.6.RP.A.3a
23	MC	1	CCSS.Math.Content.6.NS.C.6b
24	MC	1	CCSS.Math.Content.6.EE.A.2c
25	MC	1	CCSS.Math.Content.6.EE.B.5
26	MC	1	CCSS.Math.Content.6.NS.C.6c
27	MC	1	CCSS.Math.Content.6.RP.A.3c
28	MC	1	CCSS.Math.Content.6.RP.A.2
29	MC	1	CCSS.Math.Content.6.EE.B.6
30	MC	1	CCSS.Math.Content.6.NS.B.4
31	MC	1	CCSS.Math.Content.6.EE.A.3
32	MC	1	CCSS.Math.Content.6.G.A.4
33	MC	1	CCSS.Math.Content.6.NS.C.5
34	MC	1	CCSS.Math.Content.6.EE.A.3
35	MC	1	CCSS.Math.Content.6.EE.B.6
36	MC	1	CCSS.Math.Content.6.RP.A.1
37	MC	1	CCSS.Math.Content.5.G.A.1
38	MC	1	CCSS.Math.Content.6.G.A.1
39	MC	1	CCSS.Math.Content.6.NS.C.8
40	MC	1	CCSS.Math.Content.6.RP.A.3c
41	MC	1	CCSS.Math.Content.6.NS.A.1

Table G10. Mathematics Grade 6 Operational Item Map (cont.)

Item	Type	Points	Standard
42	MC	1	CCSS.Math.Content.5.OA.B.3
43	MC	1	CCSS.Math.Content.6.NS.C.8
44	MC	1	CCSS.Math.Content.5.OA.B.3
45	MC	1	CCSS.Math.Content.6.G.A.1
46	MC	1	CCSS.Math.Content.6.EE.A.2a
47	MC	1	CCSS.Math.Content.5.G.A.2
48	MC	1	CCSS.Math.Content.6.G.A.2
49	CR	2	CCSS.Math.Content.6.NS.B.4
50	CR	2	CCSS.Math.Content.6.G.A.3
51	CR	2	CCSS.Math.Content.6.RP.A.2
52	CR	2	CCSS.Math.Content.6.RP.A.3d
53	CR	2	CCSS.Math.Content.6.EE.B.7
54	CR	2	CCSS.Math.Content.6.G.A.2
55	CR	3	CCSS.Math.Content.6.EE.A.4
56	CR	3	CCSS.Math.Content.6.EE.C.9
57	CR	3	CCSS.Math.Content.6.RP.A.3b
58	CR	3	CCSS.Math.Content.6.RP.A.3b

Table G11. Mathematics Grade 7 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.7.NS.A.1d
2	MC	1	CCSS.Math.Content.7.EE.B.4b
3	MC	1	CCSS.Math.Content.7.RP.A.1
4	MC	1	CCSS.Math.Content.7.SP.B.4
5	MC	1	CCSS.Math.Content.7.EE.A.1
6	MC	1	CCSS.Math.Content.7.RP.A.3
7	MC	1	CCSS.Math.Content.7.NS.A.3
8	MC	1	CCSS.Math.Content.7.EE.A.1
9	MC	1	CCSS.Math.Content.7.RP.A.2d
10	MC	1	CCSS.Math.Content.7.NS.A.3
11	MC	1	CCSS.Math.Content.7.EE.A.2
12	MC	1	CCSS.Math.Content.7.EE.B.4a
13	MC	1	CCSS.Math.Content.7.SP.A.2
14	MC	1	CCSS.Math.Content.7.RP.A.2b
15	MC	1	CCSS.Math.Content.7.RP.A.1
16	MC	1	CCSS.Math.Content.7.SP.B.3
17	MC	1	CCSS.Math.Content.7.G.A.1
18	MC	1	CCSS.Math.Content.7.NS.A.2a
19	MC	1	CCSS.Math.Content.7.SP.C.8b
20	MC	1	CCSS.Math.Content.7.SP.C.6

Table G11. Mathematics Grade 7 Operational Item Map (cont.)

Item	Type	Points	Standard
21	MC	1	CCSS.Math.Content.7.NS.A.3
22	MC	1	CCSS.Math.Content.7.EE.A.1
23	MC	1	CCSS.Math.Content.7.EE.B.4b
24	MC	1	CCSS.Math.Content.7.EE.B.4a
25	MC	1	CCSS.Math.Content.7.G.B.4
26	MC	1	CCSS.Math.Content.7.EE.B.4a
27	MC	1	CCSS.Math.Content.7.SP.C.6
28	MC	1	CCSS.Math.Content.7.RP.A.2a
29	MC	1	CCSS.Math.Content.7.EE.B.3
30	MC	1	CCSS.Math.Content.6.SP.A.2
31	MC	1	CCSS.Math.Content.7.RP.A.3
32	MC	1	CCSS.Math.Content.7.EE.A.2
33	MC	1	CCSS.Math.Content.7.RP.A.3
34	MC	1	CCSS.Math.Content.7.EE.A.2
35	MC	1	CCSS.Math.Content.7.RP.A.2c
36	MC	1	CCSS.Math.Content.7.EE.B.4b
37	MC	1	CCSS.Math.Content.7.RP.A.1
38	MC	1	CCSS.Math.Content.7.G.A.1
39	MC	1	CCSS.Math.Content.7.RP.A.3
40	MC	1	CCSS.Math.Content.7.EE.A.1
41	MC	1	CCSS.Math.Content.7.SP.C.6
42	MC	1	CCSS.Math.Content.7.SP.C.5
43	MC	1	CCSS.Math.Content.7.SP.A.1
44	MC	1	CCSS.Math.Content.7.EE.A.1
45	MC	1	CCSS.Math.Content.7.RP.A.2c
46	MC	1	CCSS.Math.Content.7.EE.B.3
47	MC	1	CCSS.Math.Content.7.RP.A.3
48	MC	1	CCSS.Math.Content.7.SP.A.2
49	CR	2	CCSS.Math.Content.7.NS.A.1d
50	CR	2	CCSS.Math.Content.7.RP.A.3
51	CR	2	CCSS.Math.Content.7.G.B.4
52	CR	2	CCSS.Math.Content.7.NS.A.3
53	CR	2	CCSS.Math.Content.7.RP.A.2b
54	CR	2	CCSS.Math.Content.7.NS.A.2d
55	CR	3	CCSS.Math.Content.7.EE.B.3
56	CR	3	CCSS.Math.Content.7.NS.A.3
57	CR	3	CCSS.Math.Content.7.RP.A.3
58	CR	3	CCSS.Math.Content.7.EE.B.4a

Table G12. Mathematics Grade 8 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.8.EE.A.3
2	MC	1	CCSS.Math.Content.8.EE.A.1
3	MC	1	CCSS.Math.Content.8.G.A.3
4	MC	1	CCSS.Math.Content.8.F.A.3
5	MC	1	CCSS.Math.Content.8.G.A.2
6	MC	1	CCSS.Math.Content.8.G.A.2
7	MC	1	CCSS.Math.Content.8.EE.A.3
8	MC	1	CCSS.Math.Content.8.G.A.5
9	MC	1	CCSS.Math.Content.8.EE.A.4
10	MC	1	CCSS.Math.Content.8.SP.A.2
11	MC	1	CCSS.Math.Content.8.EE.A.4
12	MC	1	CCSS.Math.Content.8.EE.A.4
13	MC	1	CCSS.Math.Content.8.G.A.4
14	MC	1	CCSS.Math.Content.8.F.B.4
15	MC	1	CCSS.Math.Content.8.EE.C.7b
16	MC	1	CCSS.Math.Content.8.EE.C.8a
17	MC	1	CCSS.Math.Content.8.F.A.1
18	MC	1	CCSS.Math.Content.8.EE.A.1
19	MC	1	CCSS.Math.Content.8.F.A.2
20	MC	1	CCSS.Math.Content.7.G.B.5
21	MC	1	CCSS.Math.Content.8.SP.A.1
22	MC	1	CCSS.Math.Content.8.EE.A.1
23	MC	1	CCSS.Math.Content.8.F.A.2
24	MC	1	CCSS.Math.Content.8.EE.C.7a
25	MC	1	CCSS.Math.Content.8.G.A.1
26	MC	1	CCSS.Math.Content.8.EE.C.8b
27	MC	1	CCSS.Math.Content.8.F.B.5
28	MC	1	CCSS.Math.Content.8.EE.B.6
29	MC	1	CCSS.Math.Content.7.G.A.2
30	MC	1	CCSS.Math.Content.8.EE.B.6
31	MC	1	CCSS.Math.Content.8.SP.A.3
32	MC	1	CCSS.Math.Content.8.EE.C.8b
33	MC	1	CCSS.Math.Content.8.EE.B.6
34	MC	1	CCSS.Math.Content.8.SP.A.4
35	MC	1	CCSS.Math.Content.8.G.A.3
36	MC	1	CCSS.Math.Content.8.F.A.2
37	MC	1	CCSS.Math.Content.7.G.A.3
38	MC	1	CCSS.Math.Content.8.F.A.3
39	MC	1	CCSS.Math.Content.8.F.B.5
40	MC	1	CCSS.Math.Content.8.F.A.2
41	MC	1	CCSS.Math.Content.8.F.B.4

Table G12. Mathematics Grade 8 Operational Item Map (cont.)

Item	Type	Points	Standard
42	MC	1	CCSS.Math.Content.8.G.A.2
43	MC	1	CCSS.Math.Content.8.F.A.1
44	MC	1	CCSS.Math.Content.8.F.B.4
45	MC	1	CCSS.Math.Content.8.G.A.3
46	MC	1	CCSS.Math.Content.8.F.A.3
47	MC	1	CCSS.Math.Content.8.SP.A.3
48	MC	1	CCSS.Math.Content.8.SP.A.2
49	CR	2	CCSS.Math.Content.8.EE.B.5
50	CR	2	CCSS.Math.Content.8.G.C.9
51	CR	2	CCSS.Math.Content.8.EE.C.8a
52	CR	2	CCSS.Math.Content.8.F.A.3
53	CR	2	CCSS.Math.Content.8.G.C.9
54	CR	2	CCSS.Math.Content.8.EE.C.7b
55	CR	3	CCSS.Math.Content.8.F.B.4
56	CR	3	CCSS.Math.Content.8.EE.B.5
57	CR	3	CCSS.Math.Content.8.G.A.4
58	CR	3	CCSS.Math.Content.8.EE.C.8c

Appendix H: ELA Short-Response Rubric

2-Point Rubric–Short Response

Score	Response Features
2 Point	<p>The features of a 2-point response are</p> <ul style="list-style-type: none"> • Valid inferences and/or claims from the text where required by the prompt • Evidence of analysis of the text where required by the prompt • Relevant facts, definitions, concrete details, and/or other information from the text to develop response according to the requirements of the prompt • Sufficient number of facts, definitions, concrete details, and/or other information from the text as required by the prompt • Complete sentences where errors do not impact readability
1 Point	<p>The features of a 1-point response are</p> <ul style="list-style-type: none"> • A mostly literal recounting of events or details from the text as required by the prompt • Some relevant facts, definitions, concrete details, and/or other information from the text to develop response according to the requirements of the prompt • Incomplete sentences or bullets
0 Point*	<p>The features of a 0-point response are</p> <ul style="list-style-type: none"> • A response that does not address any of the requirements of the prompt or is totally inaccurate • A response that is not written in English • A response that is unintelligible or indecipherable

* Condition Code A is applied whenever a student who is present or test session leaves an entire constructed-response question in that session completely blank (no response attempted).

- If the prompt requires two texts and the student only references one text, the response can be scored no higher than a 1.

Appendix I: ELA Extended-Response Rubric

New York State Grade 3 Expository Writing Evaluation Rubric

CRITERIA	CCLS	SCORE				
		4 Essays at this level:	3 Essays at this level:	2 Essays at this level:	1 Essays at this level:	0* Essays at this level:
CONTENT AND ANALYSIS: the extent to which the essay conveys ideas and information clearly and accurately in order to support analysis of topics or text	W.2, R.1–9	–clearly introduce a topic in a manner that follows logically from the task and purpose –demonstrate comprehension and analysis of the text	–clearly introduce a topic in a manner that follows from the task and purpose –demonstrate grade-appropriate comprehension of the text	–introduce a topic in a manner that follows generally from the task and purpose –demonstrate a confused comprehension of the text	–introduce a topic in a manner that does not logically follow from the task and purpose –demonstrate little understanding of the text	–demonstrate a lack of comprehension of the text or task
COMMAND OF EVIDENCE: the extent to which the essay presents evidence from the provided text to support analysis and reflection	W.2 R.1–8	–develop the topic with relevant, well-chosen facts, definitions, and details throughout the essay	–develop the topic with relevant facts, definitions, and details throughout the essay	–partially develop the topic of the essay with the use of some textual evidence, some of which may be irrelevant	–demonstrate an attempt to use evidence, but only develop ideas with minimal, occasional evidence which is generally invalid or irrelevant	–provide no evidence or provide evidence that is completely irrelevant
COHERENCE, ORGANIZATION, AND STYLE: the extent to which the essay logically organizes complex ideas, concepts, and information using formal style and precise language	W.2 L.3 L.6	–clearly and consistently group related information together –skillfully connect ideas within categories of information using linking words and phrases –provide a concluding statement that follows clearly from the topic and information presented	–generally group related information together –connect ideas within categories of information using linking words and phrases –provide a concluding statement that follows from the topic and information presented	–exhibit some attempt to group related information together –inconsistently connect ideas using some linking words and phrases –provide a concluding statement that follows generally from the topic and information presented	–exhibit little attempt at organization –lack the use of linking words and phrases –provide a concluding statement that is illogical or unrelated to the topic and information presented	–exhibit no evidence of organization –do not provide a concluding statement
CONTROL OF CONVENTIONS: the extent to which the essay demonstrates command of the conventions of standard English grammar, usage, capitalization, punctuation, and spelling	W.2 L.1 L.2	–demonstrate grade-appropriate command of conventions, with few errors	–demonstrate grade-appropriate command of conventions, with occasional errors that do not hinder comprehension	–demonstrate emerging command of conventions, with some errors that may hinder comprehension	–demonstrate a lack of command of conventions, with frequent errors that hinder comprehension	–are minimal, making assessment of conventions unreliable

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

- If the student writes only a personal response and makes no reference to the text(s), the response can be scored no higher than a 1.
- Responses totally unrelated to the topic, illegible, or incoherent should be given a 0.
- A response totally copied from the text(s) with no original student writing should be scored a 0.

New York State Grade 4-5 Expository Writing Evaluation Rubric

CRITERIA	CCLS	SCORE				
		4 Essays at this level:	3 Essays at this level:	2 Essays at this level:	1 Essays at this level:	0* Essays at this level:
CONTENT AND ANALYSIS: the extent to which the essay conveys ideas and information clearly and accurately in order to support an analysis of topics or texts	W.2 R.1–9	– clearly introduce a topic in a manner that follows logically from the task and purpose – demonstrate insightful comprehension and analysis of the text(s)	– clearly introduce a topic in a manner that follows from the task and purpose – demonstrate grade-appropriate comprehension and analysis of the text(s)	– introduce a topic in a manner that follows generally from the task and purpose – demonstrate a literal comprehension of the text(s)	– introduce a topic in a manner that does not logically follow from the task and purpose – demonstrate little understanding of the text(s)	– demonstrate a lack of comprehension of the text(s) or task
COMMAND OF EVIDENCE: the extent to which the essay presents evidence from the provided texts to support analysis and reflection	W.2 W.9 R.1–9	– develop the topic with relevant, well-chosen facts, definitions, concrete details, quotations, or other information and examples from the text(s) – sustain the use of varied, relevant evidence	– develop the topic with relevant facts, definitions, details, quotations, or other information and examples from the text(s) – sustain the use of relevant evidence, with some lack of variety	– partially develop the topic of the essay with the use of some textual evidence, some of which may be irrelevant – use relevant evidence with inconsistency	– demonstrate an attempt to use evidence, but only develop ideas with minimal, occasional evidence which is generally invalid or irrelevant	– provide no evidence or provide evidence that is completely irrelevant
COHERENCE, ORGANIZATION, AND STYLE: the extent to which the essay logically organizes complex ideas, concepts, and information using formal style and precise language	W.2 L.3 L.6	– exhibit clear, purposeful organization – skillfully link ideas using grade-appropriate words and phrases – use grade-appropriate, stylistically sophisticated language and domain-specific vocabulary – provide a concluding statement that follows clearly from the topic and information presented	– exhibit clear organization – link ideas using grade-appropriate words and phrases – use grade-appropriate precise language and domain-specific vocabulary – provide a concluding statement that follows from the topic and information presented	– exhibit some attempt at organization – inconsistently link ideas using words and phrases – inconsistently use appropriate language and domain-specific vocabulary – provide a concluding statement that follows generally from the topic and information presented	– exhibit little attempt at organization, or attempts to organize are irrelevant to the task – lack the use of linking words and phrases – use language that is imprecise or inappropriate for the text(s) and task – provide a concluding statement that is illogical or unrelated to the topic and information presented	– exhibit no evidence of organization – exhibit no use of linking words and phrases – use language that is predominantly incoherent or copied directly from the text(s) – do not provide a concluding statement
CONTROL OF CONVENTIONS: the extent to which the essay demonstrates command of the conventions of standard English grammar, usage, capitalization, punctuation, and spelling	W.2 L.1 L.2	– demonstrate grade-appropriate command of conventions, with few errors	– demonstrate grade-appropriate command of conventions, with occasional errors that do not hinder comprehension	– demonstrate emerging command of conventions, with some errors that may hinder comprehension	– demonstrate a lack of command of conventions, with frequent errors that hinder comprehension	– are minimal, making assessment of conventions unreliable

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

- If the prompt requires two texts and the student only references one text, the response can be scored no higher than a 2.
- If the student writes only a personal response and makes no reference to the text(s), the response can be scored no higher than a 1.
- Responses totally unrelated to the topic, illegible, or incoherent should be given a 0.
- A response totally copied from the text(s) with no original student writing should be scored a 0.

New York State Grade 6-8 Expository Writing Evaluation Rubric

CRITERIA	CCLS	SCORE				
		4 Essays at this level:	3 Essays at this level:	2 Essays at this level:	1 Essays at this level:	0* Essays at this level:
CONTENT AND ANALYSIS: the extent to which the essay conveys complex ideas and information clearly and accurately in order to support claims in an analysis of topics or texts	W.2, R.1-9	–clearly introduce a topic in a manner that is compelling and follows logically from the task and purpose –demonstrate insightful analysis of the text(s)	– clearly introduce a topic in a manner that follows from the task and purpose –demonstrate grade-appropriate analysis of the text(s)	–introduce a topic in a manner that follows generally from the task and purpose –demonstrate a literal comprehension of the text(s)	–introduce a topic in a manner that does not logically follow from the task and purpose –demonstrate little understanding of the text(s)	–demonstrate a lack of comprehension of the text(s) or task
COMMAND OF EVIDENCE: the extent to which the essay presents evidence from the provided texts to support analysis and reflection	W.9, R.1-9	–develop the topic with relevant, well-chosen facts, definitions, concrete details, quotations, or other information and examples from the text(s) –sustain the use of varied, relevant evidence	–develop the topic with relevant facts, definitions, details, quotations, or other information and examples from the text(s) –sustain the use of relevant evidence, with some lack of variety	–partially develop the topic of the essay with the use of some textual evidence, some of which may be irrelevant –use relevant evidence with inconsistency	–demonstrate an attempt to use evidence, but only develop ideas with minimal, occasional evidence which is generally invalid or irrelevant	–provide no evidence or provide evidence that is completely irrelevant
COHERENCE, ORGANIZATION, AND STYLE: the extent to which the essay logically organizes complex ideas, concepts, and information using formal style and precise language	W.2, L.3, L.6	–exhibit clear organization, with the skillful use of appropriate and varied transitions to create a unified whole and enhance meaning –establish and maintain a formal style, using grade-appropriate, stylistically sophisticated language and domain-specific vocabulary with a notable sense of voice –provide a concluding statement or section that is compelling and follows clearly from the topic and information presented	–exhibit clear organization, with the use of appropriate transitions to create a unified whole –establish and maintain a formal style using precise language and domain-specific vocabulary –provide a concluding statement or section that follows from the topic and information presented	–exhibit some attempt at organization, with inconsistent use of transitions –establish but fail to maintain a formal style, with inconsistent use of language and domain-specific vocabulary –provide a concluding statement or section that follows generally from the topic and information presented	–exhibit little attempt at organization, or attempts to organize are irrelevant to the task –lack a formal style, using language that is imprecise or inappropriate for the text(s) and task –provide a concluding statement or section that is illogical or unrelated to the topic and information presented	–exhibit no evidence of organization –use language that is predominantly incoherent or copied directly from the text(s) –do not provide a concluding statement or section
CONTROL OF CONVENTIONS: the extent to which the essay demonstrates command of the conventions of standard English grammar, usage, capitalization, punctuation, and spelling	W.2, L.1, L.2	–demonstrate grade-appropriate command of conventions, with few errors	–demonstrate grade-appropriate command of conventions, with occasional errors that do not hinder comprehension	–demonstrate emerging command of conventions, with some errors that may hinder comprehension	–demonstrate a lack of command of conventions, with frequent errors that hinder comprehension	–are minimal, making assessment of conventions unreliable

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

- If the prompt requires two texts and the student only references one text, the response can be scored no higher than a 2.
 - If the student writes only a personal response and makes no reference to the text(s), the response can be scored no higher than a 1.
- Responses totally unrelated to the topic, illegible, or incoherent should be given a 0.
- A response totally copied from the text(s) with no original student writing should be scored a 0.

Appendix J: Mathematics Short-Response Rubric

2-Point Holistic Rubric

2 Points	<p>A two-point response includes the correct solution to the question and demonstrates a thorough understanding of the mathematical concepts and/or procedures in the task.</p> <p>This response</p> <ul style="list-style-type: none">• indicates that the student has completed the task correctly, using mathematically sound procedures• contains sufficient work to demonstrate a thorough understanding of the mathematical concepts and/or procedures• may contain inconsequential errors that do not detract from the correct solution and the demonstration of a thorough understanding
1 Point	<p>A one-point response demonstrates only a partial understanding of the mathematical concepts and/or procedures in the task.</p> <p>This response</p> <ul style="list-style-type: none">• correctly addresses only some elements of the task• may contain an incorrect solution but applies a mathematically appropriate process• may contain the correct solution but required work is incomplete
0 Points*	<p>A zero-point response is incorrect, irrelevant, incoherent, or contains a correct solution obtained using an obviously incorrect procedure. Although some elements may contain correct mathematical procedures, holistically they are not sufficient to demonstrate even a limited understanding of the mathematical concepts embodied in the task.</p>

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

Appendix K: Mathematics Extended-Response Rubric

3-Point Holistic Rubric

3 Points	<p>A three-point response includes the correct solution(s) to the question and demonstrates a thorough understanding of the mathematical concepts and/or procedures in the task.</p> <p>This response</p> <ul style="list-style-type: none"> • indicates that the student has completed the task correctly, using mathematically sound procedures • contains sufficient work to demonstrate a thorough understanding of the mathematical concepts and/or procedures • may contain inconsequential errors that do not detract from the correct solution(s) and the demonstration of a thorough understanding
2 Points	<p>A two-point response demonstrates a partial understanding of the mathematical concepts and/or procedures in the task.</p> <p>This response</p> <ul style="list-style-type: none"> • appropriately addresses most, but not all, aspects of the task using mathematically sound procedures • may contain an incorrect solution but provides sound procedures, reasoning, and/or explanations • may reflect some minor misunderstanding of the underlying mathematical concepts and/or procedures
1 Point	<p>A one-point response demonstrates only a limited understanding of the mathematical concepts and/or procedures in the task.</p> <p>This response</p> <ul style="list-style-type: none"> • may address some elements of the task correctly but reaches an inadequate solution and/or provides reasoning that is faulty or incomplete • exhibits multiple flaws related to misunderstanding of important aspects of the task, misuse of mathematical procedures, or faulty mathematical reasoning • reflects a lack of essential understanding of the underlying mathematical concepts • may contain the correct solution(s) but required work is limited
0 Points*	<p>A zero-point response is incorrect, irrelevant, incoherent, or contains a correct solution obtained using an obviously incorrect procedure. Although some elements may contain correct mathematical procedures, holistically they are not sufficient to demonstrate even a limited understanding of the mathematical concepts embodied in the task.</p>

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

Appendix L: Factor Analysis Results for Select Subgroups

As described in Section 3, “Validity,” a principal components factor analysis was conducted on the Grades 3–8 Common Core ELA and Mathematics Tests data. The analyses were conducted for the total population of students and select subgroups: ELL, SWD, SUA, SWD students using disability accommodations (SWD & SUA), and ELL students using ELL-related accommodations (ELL & SUA). Tables L1 and L2 contain the results of factor analysis on the subpopulation data for the Grades 3–8 Common Core ELA and Mathematics Tests, respectively.

Table L1. ELA Grade 3 Test Factor Analyses for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	1	5.24	13.80	13.80
		2	1.48	3.88	17.68
		3	1.16	3.04	20.72
		4	1.08	2.84	23.56
		5	1.07	2.82	26.38
		6	1.06	2.79	29.17
		7	1.05	2.76	31.93
		8	1.02	2.67	34.60
		9	1.00	2.64	37.24
SWD	All Codes	1	6.14	16.15	16.15
		2	1.50	3.94	20.09
		3	1.14	3.00	23.09
		4	1.06	2.80	25.89
		5	1.04	2.73	28.62
		6	1.02	2.68	31.30
		7	1.02	2.67	33.97
		8	1.01	2.64	36.61
SUA	All Codes	1	6.13	16.12	16.12
		2	1.50	3.96	20.08
		3	1.13	2.98	23.06
		4	1.07	2.81	25.87
		5	1.03	2.70	28.57
		6	1.02	2.69	31.26
		7	1.01	2.66	33.92

Table L1. ELA Grade 3 Test Factor Analyses for Select Subgroups (cont.)

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SWD/SUA	SUA=504 plan codes	1	5.60	14.74	14.74
		2	1.49	3.92	18.66
		3	1.13	2.97	21.63
		4	1.10	2.90	24.53
		5	1.06	2.79	27.32
		6	1.04	2.74	30.06
		7	1.03	2.72	32.78
		8	1.02	2.69	35.47
		9	1.01	2.67	38.14
ELL/SUA	SUA & ELL codes	1	5.11	13.44	13.44
		2	1.44	3.78	17.22
		3	1.17	3.09	20.31
		4	1.12	2.94	23.25
		5	1.09	2.86	26.11
		6	1.08	2.84	28.95
		7	1.05	2.77	31.72
		8	1.03	2.72	34.44
		9	1.02	2.69	37.13

Table L2. ELA Grade 4 Test Factor Analyses for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	1	6.19	15.09	15.09
		2	1.64	4.00	19.09
		3	1.11	2.72	21.81
		4	1.10	2.69	24.50
		5	1.09	2.65	27.15
		6	1.07	2.60	29.75
		7	1.06	2.59	32.34
		8	1.03	2.52	34.86
		9	1.02	2.49	37.35

Table L2. ELA Grade 4 Test Factor Analyses for Select Subgroups (cont.)

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SWD	All Codes	1	7.16	17.47	17.47
		2	1.63	3.97	21.44
		3	1.10	2.69	24.13
		4	1.07	2.62	26.75
		5	1.05	2.57	29.32
		6	1.05	2.56	31.88
		7	1.04	2.53	34.41
		8	1.03	2.50	36.91
		9	1.01	2.45	39.36
SUA	All Codes	1	7.09	17.30	17.30
		2	1.60	3.89	21.19
		3	1.09	2.67	23.86
		4	1.07	2.60	26.46
		5	1.05	2.57	29.03
		6	1.05	2.55	31.58
		7	1.03	2.50	34.08
		8	1.02	2.48	36.56
		9	1.01	2.47	39.03
SWD/SUA	SUA=504 plan codes	1	6.58	16.05	16.05
		2	1.59	3.88	19.93
		3	1.11	2.70	22.63
		4	1.10	2.67	25.30
		5	1.08	2.64	27.94
		6	1.07	2.61	30.55
		7	1.05	2.56	33.11
		8	1.04	2.53	35.64
		9	1.02	2.49	38.13
ELL/SUA	SUA & ELL codes	1	5.82	14.21	14.21
		2	1.46	3.55	17.76
		3	1.14	2.78	20.54
		4	1.13	2.76	23.30
		5	1.10	2.69	25.99
		6	1.09	2.66	28.65
		7	1.08	2.64	31.29
		8	1.06	2.58	33.87
		9	1.04	2.54	36.41

Table L3. ELA Grade 5 Test Factor Analyses for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	1	6.65	12.78	12.78
		2	1.60	3.07	15.85
		3	1.27	2.44	18.29
		4	1.13	2.18	20.47
		5	1.10	2.11	22.58
		6	1.09	2.10	24.68
		7	1.08	2.08	26.76
		8	1.06	2.05	28.81
		9	1.05	2.03	30.84
SWD	All Codes	1	7.75	14.91	14.91
		2	1.70	3.28	18.19
		3	1.23	2.37	20.56
		4	1.12	2.16	22.72
		5	1.07	2.06	24.78
		6	1.06	2.04	26.82
		7	1.05	2.02	28.84
		8	1.02	1.96	30.80
		9	1.01	1.94	32.74
SUA	All Codes	1	7.77	14.94	14.94
		2	1.69	3.25	18.19
		3	1.22	2.35	20.54
		4	1.11	2.14	22.68
		5	1.08	2.08	24.76
		6	1.06	2.05	26.81
		7	1.05	2.02	28.83
		8	1.02	1.96	30.79
		9	1.01	1.95	32.74
SWD/SUA	SUA=504 plan codes	1	7.04	13.53	13.53
		2	1.68	3.22	16.75
		3	1.21	2.33	19.08
		4	1.14	2.19	21.27
		5	1.10	2.11	23.38
		6	1.09	2.10	25.48
		7	1.07	2.06	27.54
		8	1.05	2.01	29.55
		9	1.03	1.98	31.53

Table L3. ELA Grade 5 Test Factor Analyses for Select Subgroups (cont.)

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL/SUA	SUA & ELL codes	1	6.01	11.56	11.56
		2	1.51	2.91	14.47
		3	1.28	2.46	16.93
		4	1.20	2.31	19.24
		5	1.15	2.21	21.45
		6	1.13	2.17	23.62
		7	1.12	2.16	25.78
		8	1.11	2.13	27.91
		9	1.08	2.08	29.99

Table L4. ELA Grade 6 Test Factor Analyses for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	1	6.74	12.96	12.96
		2	1.91	3.67	16.63
		3	1.25	2.40	19.03
		4	1.17	2.25	21.28
		5	1.12	2.16	23.44
		6	1.08	2.09	25.53
		7	1.08	2.07	27.60
		8	1.06	2.04	29.64
		9	1.05	2.02	31.66
SWD	All Codes	1	8.18	15.73	15.73
		2	2.01	3.86	19.59
		3	1.28	2.46	22.05
		4	1.15	2.20	24.25
		5	1.05	2.02	26.27
		6	1.05	2.02	28.29
		7	1.04	2.01	30.30
		8	1.03	1.97	32.27

Table L4. ELA Grade 6 Test Factor Analyses for Select Subgroups (cont.)

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SUA	All Codes	1	8.67	16.68	16.68
		2	1.97	3.79	20.47
		3	1.29	2.48	22.95
		4	1.13	2.18	25.13
		5	1.04	2.01	27.14
		6	1.04	2.00	29.14
		7	1.02	1.97	31.11
		8	1.01	1.94	33.05
SWD/SUA	SUA=504 plan codes	1	7.91	15.21	15.21
		2	1.98	3.80	19.01
		3	1.27	2.44	21.45
		4	1.15	2.22	23.67
		5	1.08	2.07	25.74
		6	1.06	2.03	27.77
		7	1.05	2.01	29.78
		8	1.03	1.99	31.77
		9	1.01	1.94	33.71
ELL/SUA	SUA & ELL codes	1	6.77	13.02	13.02
		2	1.80	3.46	16.48
		3	1.25	2.41	18.89
		4	1.21	2.32	21.21
		5	1.14	2.20	23.41
		6	1.12	2.15	25.56
		7	1.10	2.11	27.67
		8	1.09	2.09	29.76
		9	1.07	2.05	31.81

Table L5. ELA Grade 7 Test Factor Analyses for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	1	5.97	11.70	11.70
		2	1.71	3.36	15.06
		3	1.17	2.30	17.36
		4	1.16	2.27	19.63
		5	1.13	2.21	21.84
		6	1.11	2.17	24.01
		7	1.10	2.16	26.17
		8	1.10	2.16	28.33
		9	1.08	2.12	30.45
SWD	All Codes	1	6.95	13.62	13.62
		2	1.84	3.61	17.23
		3	1.21	2.37	19.60
		4	1.15	2.26	21.86
		5	1.09	2.13	23.99
		6	1.06	2.08	26.07
		7	1.06	2.08	28.15
		8	1.04	2.03	30.18
		9	1.03	2.02	32.20
SUA	All Codes	1	7.40	14.51	14.51
		2	1.82	3.56	18.07
		3	1.21	2.38	20.45
		4	1.11	2.19	22.64
		5	1.08	2.11	24.75
		6	1.05	2.06	26.81
		7	1.04	2.04	28.85
		8	1.04	2.04	30.89
		9	1.03	2.03	32.92
SWD/SUA	SUA=504 plan codes	1	6.52	12.79	12.79
		2	1.81	3.55	16.34
		3	1.22	2.39	18.73
		4	1.14	2.23	20.96
		5	1.10	2.15	23.11
		6	1.08	2.13	25.24
		7	1.07	2.09	27.33
		8	1.06	2.08	29.41
		9	1.05	2.06	31.47

Table L5. ELA Grade 7 Test Factor Analyses for Select Subgroups (cont.)

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL/SUA	SUA & ELL codes	1	5.42	10.64	10.64
		2	1.60	3.14	13.78
		3	1.24	2.44	16.22
		4	1.20	2.35	18.57
		5	1.19	2.34	20.91
		6	1.17	2.29	23.20
		7	1.15	2.25	25.45
		8	1.14	2.23	27.68
		9	1.12	2.19	29.87

Table L6. ELA Grade 8 Test Factor Analyses for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	1	6.00	11.55	11.55
		2	2.02	3.89	15.44
		3	1.31	2.51	17.95
		4	1.15	2.20	20.15
		5	1.12	2.16	22.31
		6	1.12	2.15	24.46
		7	1.10	2.12	26.58
		8	1.09	2.09	28.67
		9	1.06	2.04	30.71
SWD	All Codes	1	7.23	13.91	13.91
		2	1.96	3.77	17.68
		3	1.27	2.45	20.13
		4	1.12	2.15	22.28
		5	1.07	2.06	24.34
		6	1.06	2.04	26.38
		7	1.05	2.02	28.40
		8	1.04	1.99	30.39
		9	1.03	1.97	32.36

Table L6. ELA Grade 8 Test Factor Analyses for Select Subgroups (cont.)

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SUA	All Codes	1	7.92	15.23	15.23
		2	1.95	3.76	18.99
		3	1.27	2.44	21.43
		4	1.11	2.14	23.57
		5	1.06	2.03	25.60
		6	1.04	2.01	27.61
		7	1.03	1.98	29.59
		8	1.02	1.97	31.56
		9	1.01	1.94	33.50
SWD/SUA	SUA=504 plan codes	1	7.10	13.65	13.65
		2	1.92	3.70	17.35
		3	1.27	2.44	19.79
		4	1.11	2.14	21.93
		5	1.09	2.10	24.03
		6	1.07	2.06	26.09
		7	1.07	2.06	28.15
		8	1.04	2.00	30.15
		9	1.03	1.99	32.14
ELL/SUA	SUA & ELL codes	1	5.51	10.60	10.60
		2	1.83	3.52	14.12
		3	1.33	2.56	16.68
		4	1.25	2.40	19.08
		5	1.19	2.29	21.37
		6	1.17	2.26	23.63
		7	1.15	2.21	25.84
		8	1.13	2.17	28.01
		9	1.11	2.13	30.14

Table L7. Mathematics Grade 3 Test Factor Analyses for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	1	10.28	20.99	20.99
		2	2.29	4.67	25.66
		3	1.22	2.49	28.15
		4	1.13	2.31	30.46
		5	1.04	2.12	32.58
		6	1.02	2.08	34.66
SWD	All Codes	1	11.01	22.47	22.47
		2	2.02	4.12	26.59
		3	1.25	2.54	29.13
		4	1.08	2.21	31.34
		5	1.03	2.10	33.44
SUA	All Codes	1	11.32	23.10	23.10
		2	2.02	4.13	27.23
		3	1.22	2.49	29.72
		4	1.08	2.20	31.92
		5	1.02	2.09	34.01
SWD/SUA	SUA=504 plan codes	1	10.75	21.94	21.94
		2	1.98	4.04	25.98
		3	1.27	2.59	28.57
		4	1.10	2.25	30.82
		5	1.04	2.12	32.94
		6	1.00	2.05	34.99
ELL/SUA	SUA & ELL codes	1	10.52	21.47	21.47
		2	2.08	4.25	25.72
		3	1.22	2.49	28.21
		4	1.10	2.25	30.46
		5	1.05	2.15	32.61
		6	1.02	2.09	34.70

Table L8. Mathematics Grade 4 Test Factor Analyses for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	1	12.64	24.30	24.30
		2	1.70	3.27	27.57
		3	1.24	2.38	29.95
		4	1.10	2.11	32.06
		5	1.06	2.05	34.11
		6	1.03	1.99	36.10
SWD	All Codes	1	12.70	24.42	24.42
		2	1.61	3.09	27.51
		3	1.23	2.37	29.88
		4	1.12	2.14	32.02
		5	1.09	2.10	34.12
		6	1.02	1.96	36.08
SUA	All Codes	1	13.34	25.65	25.65
		2	1.58	3.04	28.69
		3	1.24	2.39	31.08
		4	1.11	2.13	33.21
		5	1.09	2.10	35.31
		6	1.00	1.93	37.24
SWD/SUA	SUA=504 plan codes	1	12.16	23.39	23.39
		2	1.60	3.08	26.47
		3	1.24	2.39	28.86
		4	1.14	2.20	31.06
		5	1.12	2.15	33.21
		6	1.03	1.99	35.20
ELL/SUA	SUA & ELL codes	1	12.95	24.90	24.90
		2	1.64	3.14	28.04
		3	1.26	2.42	30.46
		4	1.10	2.11	32.57
		5	1.06	2.03	34.60
		6	1.04	2.00	36.60

Table L9. Mathematics Grade 5 Test Factor Analyses for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	1	10.24	19.69	19.69
		2	2.04	3.92	23.61
		3	1.34	2.58	26.19
		4	1.12	2.16	28.35
		5	1.06	2.04	30.39
		6	1.04	2.00	32.39
		7	1.01	1.95	34.34
SWD	All Codes	1	10.10	19.43	19.43
		2	1.87	3.60	23.03
		3	1.36	2.62	25.65
		4	1.13	2.17	27.82
		5	1.05	2.03	29.85
		6	1.03	1.98	31.83
SUA	All Codes	1	10.64	20.46	20.46
		2	1.88	3.61	24.07
		3	1.37	2.63	26.70
		4	1.11	2.14	28.84
		5	1.05	2.03	30.87
		6	1.02	1.96	32.83
SWD/SUA	SUA=504 plan codes	1	9.43	18.14	18.14
		2	1.79	3.45	21.59
		3	1.39	2.68	24.27
		4	1.15	2.22	26.49
		5	1.08	2.08	28.57
		6	1.05	2.02	30.59
		7	1.02	1.97	32.56
		8	1.00	1.93	34.49
ELL/SUA	SUA & ELL codes	1	9.76	18.77	18.77
		2	1.98	3.80	22.57
		3	1.39	2.67	25.24
		4	1.15	2.22	27.46
		5	1.09	2.10	29.56
		6	1.07	2.05	31.61
		7	1.02	1.97	33.58
		8	1.00	1.93	35.51

Table L10. Mathematics Grade 6 Test Factor Analyses for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	1	10.96	18.89	18.89
		2	1.93	3.33	22.22
		3	1.29	2.22	24.44
		4	1.18	2.03	26.47
		5	1.08	1.87	28.34
		6	1.05	1.81	30.15
		7	1.04	1.79	31.94
		8	1.00	1.73	33.67
SWD	All Codes	1	10.57	18.22	18.22
		2	1.78	3.07	21.29
		3	1.31	2.26	23.55
		4	1.13	1.95	25.50
		5	1.08	1.87	27.37
		6	1.07	1.84	29.21
		7	1.02	1.75	30.96
SUA	All Codes	1	11.68	20.14	20.14
		2	1.84	3.18	23.32
		3	1.28	2.21	25.53
		4	1.12	1.94	27.47
		5	1.08	1.86	29.33
		6	1.04	1.78	31.11
SWD/SUA	SUA=504 plan codes	1	10.30	17.77	17.77
		2	1.76	3.03	20.80
		3	1.30	2.23	23.03
		4	1.14	1.96	24.99
		5	1.09	1.88	26.87
		6	1.07	1.84	28.71
		7	1.04	1.79	30.50
		8	1.02	1.75	32.25
ELL/SUA	SUA & ELL codes	1	11.37	19.61	19.61
		2	1.89	3.26	22.87
		3	1.23	2.12	24.99
		4	1.17	2.01	27.00
		5	1.08	1.86	28.86
		6	1.06	1.83	30.69
		7	1.03	1.77	32.46
		8	1.01	1.74	34.20
		9	1.01	1.74	35.94

Table L11. Mathematics Grade 7 Test Factor Analyses for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	1	10.26	17.69	17.69
		2	1.81	3.12	20.81
		3	1.26	2.17	22.98
		4	1.11	1.92	24.90
		5	1.08	1.86	26.76
		6	1.06	1.84	28.60
		7	1.03	1.77	30.37
		8	1.02	1.75	32.12
		9	1.01	1.73	33.85
SWD	All Codes	1	9.02	15.55	15.55
		2	1.68	2.90	18.45
		3	1.33	2.30	20.75
		4	1.10	1.89	22.64
		5	1.05	1.81	24.45
		6	1.05	1.81	26.26
		7	1.03	1.77	28.03
		8	1.03	1.77	29.80
		9	1.00	1.73	31.53
SUA	All Codes	1	10.10	17.41	17.41
		2	1.78	3.06	20.47
		3	1.32	2.27	22.74
		4	1.08	1.86	24.60
		5	1.05	1.81	26.41
		6	1.04	1.79	28.20
		7	1.02	1.76	29.96
		8	1.01	1.74	31.70
SWD/SUA	SUA=504 plan codes	1	8.63	14.88	14.88
		2	1.64	2.83	17.71
		3	1.35	2.33	20.04
		4	1.11	1.92	21.96
		5	1.07	1.84	23.80
		6	1.06	1.83	25.63
		7	1.05	1.81	27.44
		8	1.04	1.80	29.24
		9	1.02	1.75	30.99

Table L11. Mathematics Grade 7 Test Factor Analyses for Select Subgroups (cont.)

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL/SUA	SUA & ELL codes	1	10.00	17.25	17.25
		2	1.85	3.20	20.45
		3	1.25	2.15	22.60
		4	1.14	1.96	24.56
		5	1.11	1.91	26.47
		6	1.09	1.88	28.35
		7	1.06	1.83	30.18
		8	1.04	1.80	31.98
		9	1.03	1.78	33.76

Table L12. Mathematics Grade 8 Test Factor Analyses for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	1	10.80	18.62	18.62
		2	1.57	2.70	21.32
		3	1.19	2.05	23.37
		4	1.14	1.97	25.34
		5	1.09	1.87	27.21
		6	1.06	1.82	29.03
		7	1.02	1.76	30.79
		8	1.02	1.75	32.54
		9	1.01	1.74	34.28
SWD	All Codes	1	8.56	14.76	14.76
		2	1.51	2.61	17.37
		3	1.23	2.13	19.50
		4	1.08	1.87	21.37
		5	1.06	1.83	23.20
		6	1.05	1.81	25.01
		7	1.04	1.78	26.79
		8	1.02	1.77	28.56
		9	1.02	1.76	30.32

Table L12. Mathematics Grade 8 Test Factor Analyses for Select Subgroups (cont.)

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SUA	All Codes	1	9.48	16.35	16.35
		2	1.56	2.69	19.04
		3	1.22	2.11	21.15
		4	1.07	1.84	22.99
		5	1.05	1.81	24.80
		6	1.03	1.77	26.57
		7	1.02	1.75	28.32
		8	1.01	1.75	30.07
SWD/SUA	SUA=504 plan codes	1	8.25	14.23	14.23
		2	1.51	2.60	16.83
		3	1.25	2.16	18.99
		4	1.10	1.90	20.89
		5	1.07	1.84	22.73
		6	1.06	1.83	24.56
		7	1.05	1.80	26.36
		8	1.03	1.78	28.14
		9	1.02	1.76	29.90
ELL/SUA	SUA & ELL codes	1	9.78	16.86	16.86
		2	1.59	2.75	19.61
		3	1.16	2.01	21.62
		4	1.16	1.99	23.61
		5	1.12	1.93	25.54
		6	1.10	1.90	27.44
		7	1.07	1.84	29.28
		8	1.05	1.82	31.10
		9	1.04	1.79	32.89

Appendix M: Classical Test Theory Statistics

These tables support the classical test theory analyses described in Section 5, “Operational Test Data Collection and Classical Analysis.” They include item type, sample size, p-value, percent of omitted responses and the point-biserial of the key. External linking and field test items (i.e., those not contributing to students’ scores) have been omitted.

Table M1. ELA Grade 3 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	198,449	.77	0.02	.44
2	MC	198,360	.81	0.06	.32
3	MC	198,304	.49	0.09	.23
4	MC	198,311	.81	0.09	.40
5	MC	198,292	.70	0.10	.41
6	MC	198,202	.45	0.14	.31
7	MC	198,262	.50	0.11	.36
8	MC	198,179	.62	0.16	.34
9	MC	198,280	.55	0.10	.39
10	MC	198,152	.49	0.17	.41
11	MC	198,109	.36	0.19	.23
12	MC	198,043	.52	0.22	.44
13	MC	197,795	.43	0.35	.32
14	MC	197,693	.50	0.40	.39
15	MC	197,664	.48	0.42	.38
16	MC	197,424	.52	0.54	.35
17	MC	197,370	.63	0.56	.50
18	MC	196,980	.49	0.76	.37
19	MC	190,984	.37	3.78	.38
20	MC	190,055	.56	4.25	.44
21	MC	189,275	.34	4.64	.33
22	MC	188,815	.49	4.87	.43
23	MC	198,374	.42	0.06	.37
24	MC	198,356	.72	0.07	.39
25	MC	198,276	.75	0.11	.42
26	MC	198,193	.42	0.15	.26
27	MC	198,284	.60	0.10	.46
28	MC	198,175	.41	0.16	.27
29	MC	197,821	.44	0.34	.39

Table M1. ELA Grade 3 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
30	CR	197,481	.65	0.51	
31	CR	195,390	.53	1.56	
32	CR	193,085	.48	2.72	
33	CR	190,706	.43	3.92	
34	CR	196,535	.50	0.98	
35	CR	191,968	.46	3.28	
36	CR	195,248	.48	1.63	
37	CR	192,459	.32	3.04	
38	CR	191,174	.54	3.68	

Table M2. ELA Grade 4 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	203,735	.64	0.03	.35
2	MC	203,733	.77	0.03	.44
3	MC	203,721	.77	0.04	.38
4	MC	203,556	.49	0.12	.28
5	MC	203,665	.70	0.07	.42
6	MC	203,656	.64	0.07	.45
7	MC	203,640	.72	0.08	.46
8	MC	203,666	.25	0.07	.23
9	MC	203,646	.73	0.08	.40
10	MC	203,586	.52	0.11	.40
11	MC	203,580	.72	0.11	.49
12	MC	203,598	.73	0.10	.45
13	MC	203,477	.48	0.16	.39
14	MC	203,452	.54	0.17	.27
15	MC	203,402	.75	0.20	.45
16	MC	203,391	.55	0.20	.28
17	MC	203,351	.49	0.22	.42
18	MC	203,252	.71	0.27	.51
19	MC	201,249	.56	1.25	.26
20	MC	201,020	.54	1.36	.35
21	MC	200,646	.51	1.55	.34
22	MC	200,409	.50	1.66	.30
23	MC	200,100	.42	1.82	.24
24	MC	199,939	.38	1.89	.22
25	MC	203,713	.58	0.04	.13
26	MC	203,702	.55	0.05	.33
27	MC	203,656	.50	0.07	.32
28	MC	203,613	.48	0.09	.41

Table M2. ELA Grade 4 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
29	MC	203,666	.27	0.07	.13
30	MC	203,643	.61	0.08	.44
31	MC	203,413	.74	0.19	.46
32	CR	202,462	.63	0.66	
33	CR	202,463	.65	0.66	
34	CR	201,077	.54	1.34	
35	CR	198,681	.48	2.51	
36	CR	203,205	.50	0.29	
37	CR	202,665	.53	0.56	
38	CR	202,872	.56	0.46	
39	CR	201,163	.47	1.29	
40	CR	200,301	.60	1.72	
41	CR	199,780	.46	1.97	

Table M3. ELA Grade 5 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	199,410	.90	0.01	.34
2	MC	199,302	.55	0.06	.29
3	MC	199,321	.65	0.05	.38
4	MC	199,210	.62	0.11	.46
5	MC	199,319	.56	0.05	.29
6	MC	199,263	.55	0.08	.51
7	MC	199,275	.67	0.08	.38
8	MC	199,222	.44	0.10	.34
9	MC	199,265	.72	0.08	.34
10	MC	199,152	.54	0.14	.40
11	MC	199,127	.41	0.15	.29
12	MC	199,178	.56	0.12	.34
13	MC	199,159	.62	0.13	.45
14	MC	199,208	.64	0.11	.45
15	MC	199,120	.57	0.15	.31
16	MC	199,190	.44	0.12	.33
17	MC	199,082	.46	0.17	.26
18	MC	199,065	.72	0.18	.52
19	MC	198,863	.45	0.28	.41
20	MC	198,990	.85	0.22	.44
21	MC	198,935	.51	0.25	.44
22	MC	198,685	.80	0.37	.44
23	MC	198,546	.48	0.44	.30
24	MC	198,458	.58	0.49	.39

Table M3. ELA Grade 5 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
25	MC	198,294	.74	0.57	.48
26	MC	198,288	.72	0.57	.40
27	MC	198,028	.29	0.70	.21
28	MC	197,946	.47	0.74	.45
29	MC	196,889	.38	1.27	.34
30	MC	196,889	.77	1.27	.49
31	MC	196,619	.52	1.41	.42
32	MC	196,392	.55	1.52	.42
33	MC	196,135	.60	1.65	.34
34	MC	195,880	.63	1.78	.42
35	MC	195,588	.54	1.93	.44
36	MC	199,387	.74	0.02	.26
37	MC	199,381	.59	0.02	.37
38	MC	199,326	.26	0.05	.10
39	MC	199,350	.65	0.04	.35
40	MC	199,372	.70	0.03	.32
41	MC	199,358	.69	0.03	.29
42	MC	199,252	.77	0.09	.31
43	CR	198,962	.72	0.23	
44	CR	198,498	.69	0.47	
45	CR	198,118	.65	0.66	
46	CR	197,041	.50	1.20	
47	CR	199,226	.67	0.10	
48	CR	197,391	.47	1.02	
49	CR	198,552	.58	0.44	
50	CR	195,934	.34	1.75	
51	CR	196,745	.61	1.34	
52	CR	195,624	.44	1.91	

Table M4. ELA Grade 6 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	200,374	.60	0.03	.46
2	MC	200,173	.52	0.13	.41
3	MC	200,361	.73	0.04	.42
4	MC	200,271	.58	0.08	.41
5	MC	200,266	.68	0.09	.47
6	MC	200,298	.58	0.07	.43
7	MC	200,306	.84	0.07	.43
8	MC	200,271	.64	0.08	.45
9	MC	200,176	.54	0.13	.43
10	MC	200,326	.92	0.06	.37
11	MC	200,249	.75	0.10	.36
12	MC	200,239	.43	0.10	.26
13	MC	200,107	.42	0.17	.30
14	MC	200,238	.70	0.10	.33
15	MC	199,947	.39	0.25	.32
16	MC	200,090	.75	0.17	.47
17	MC	200,025	.82	0.21	.51
18	MC	199,870	.56	0.28	.37
19	MC	199,893	.76	0.27	.44
20	MC	199,855	.52	0.29	.32
21	MC	199,290	.62	0.57	.46
22	MC	199,344	.65	0.55	.46
23	MC	199,256	.68	0.59	.46
24	MC	199,144	.47	0.65	.34
25	MC	198,964	.54	0.74	.45
26	MC	198,818	.45	0.81	.29
27	MC	198,735	.49	0.85	.34
28	MC	198,053	.53	1.19	.36
29	MC	197,767	.41	1.33	.39
30	MC	197,477	.49	1.48	.40
31	MC	197,283	.66	1.58	.49
32	MC	196,907	.36	1.76	.27
33	MC	197,058	.54	1.69	.56
34	MC	196,772	.59	1.83	.33
35	MC	196,620	.61	1.91	.42
36	MC	200,349	.62	0.05	.42
37	MC	200,368	.69	0.04	.41
38	MC	200,300	.53	0.07	.30
39	MC	200,330	.71	0.05	.38
40	MC	200,332	.52	0.05	.43
41	MC	200,310	.68	0.06	.37

Table M4. ELA Grade 6 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
42	MC	200,239	.69	0.10	.36
43	CR	199,925	.72	0.26	
44	CR	198,188	.66	1.12	
45	CR	198,729	.59	0.85	
46	CR	196,692	.43	1.87	
47	CR	199,671	.62	0.38	
48	CR	199,408	.59	0.51	
49	CR	199,135	.64	0.65	
50	CR	198,449	.67	0.99	
51	CR	197,401	.62	1.52	
52	CR	195,972	.48	2.23	

Table M5. ELA Grade 7 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	197,530	.81	0.04	.35
2	MC	197,322	.49	0.14	.30
3	MC	197,314	.49	0.15	.22
4	MC	197,429	.55	0.09	.35
5	MC	197,396	.59	0.10	.25
6	MC	197,516	.68	0.04	.45
7	MC	197,255	.42	0.18	.30
8	MC	197,386	.75	0.11	.27
9	MC	197,309	.28	0.15	.37
10	MC	197,346	.62	0.13	.55
11	MC	197,378	.43	0.11	.29
12	MC	197,299	.45	0.15	.30
13	MC	197,071	.35	0.27	.31
14	MC	197,332	.69	0.14	.34
15	MC	197,274	.44	0.17	.26
16	MC	197,279	.64	0.16	.50
17	MC	197,265	.45	0.17	.32
18	MC	197,277	.56	0.16	.35
19	MC	197,246	.34	0.18	.30
20	MC	196,496	.64	0.56	.42
21	MC	196,263	.60	0.68	.39
22	MC	196,157	.54	0.73	.34
23	MC	196,137	.28	0.74	.34
24	MC	195,857	.57	0.88	.38
25	MC	195,723	.75	0.95	.45
26	MC	195,520	.65	1.05	.42

Table M5. ELA Grade 7 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
27	MC	195,319	.62	1.16	.53
28	MC	194,473	.75	1.58	.47
29	MC	194,301	.50	1.67	.45
30	MC	194,038	.44	1.80	.41
31	MC	193,753	.74	1.95	.49
32	MC	193,416	.39	2.12	.29
33	MC	193,195	.43	2.23	.32
34	MC	192,988	.43	2.34	.29
35	MC	197,529	.73	0.04	.36
36	MC	197,534	.70	0.03	.42
37	MC	197,433	.59	0.09	.38
38	MC	197,481	.50	0.06	.41
39	MC	197,501	.62	0.05	.36
40	MC	197,399	.49	0.10	.25
41	MC	197,367	.45	0.12	.40
42	CR	196,842	.67	0.39	
43	CR	196,375	.68	0.62	
44	CR	195,318	.59	1.16	
45	CR	194,819	.55	1.41	
46	CR	196,897	.69	0.36	
47	CR	196,348	.70	0.64	
48	CR	195,400	.55	1.11	
49	CR	195,063	.61	1.29	
50	CR	192,604	.62	2.53	
51	CR	192,491	.49	2.59	

Table M6. ELA Grade 8 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	203,307	.91	0.03	.24
2	MC	203,209	.54	0.07	.46
3	MC	203,270	.78	0.04	.44
4	MC	203,227	.57	0.06	.34
5	MC	203,059	.43	0.15	.42
6	MC	203,107	.61	0.12	.33
7	MC	203,140	.72	0.11	.45
8	MC	203,036	.56	0.16	.31
9	MC	203,014	.63	0.17	.38
10	MC	203,010	.41	0.17	.25
11	MC	203,164	.81	0.10	.25
12	MC	203,088	.72	0.13	.36

Table M6. ELA Grade 8 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
13	MC	203,100	.59	0.13	.48
14	MC	203,099	.85	0.13	.43
15	MC	203,138	.63	0.11	.39
16	MC	203,247	.86	0.05	.43
17	MC	203,075	.40	0.14	.31
18	MC	203,144	.63	0.11	.43
19	MC	203,137	.47	0.11	.38
20	MC	203,150	.41	0.10	.30
21	MC	203,051	.44	0.15	.29
22	MC	202,605	.58	0.37	.47
23	MC	202,536	.53	0.40	.38
24	MC	202,322	.66	0.51	.47
25	MC	202,327	.55	0.51	.32
26	MC	202,196	.56	0.57	.43
27	MC	202,135	.58	0.60	.45
28	MC	202,054	.67	0.64	.41
29	MC	201,709	.58	0.81	.46
30	MC	201,608	.58	0.86	.52
31	MC	201,268	.59	1.03	.40
32	MC	201,064	.49	1.13	.31
33	MC	201,009	.63	1.16	.43
34	MC	200,833	.53	1.24	.40
35	MC	200,729	.53	1.29	.40
36	MC	203,226	.60	0.06	.34
37	MC	203,235	.81	0.06	.29
38	MC	203,238	.74	0.06	.27
39	MC	203,182	.74	0.09	.44
40	MC	203,228	.62	0.06	.38
41	MC	203,158	.53	0.10	.33
42	MC	203,090	.76	0.13	.43
43	CR	202,020	.68	0.66	
44	CR	200,511	.70	1.40	
45	CR	201,360	.65	0.98	
46	CR	199,243	.52	2.02	
47	CR	202,582	.70	0.38	
48	CR	202,353	.78	0.49	
49	CR	202,076	.75	0.63	
50	CR	199,892	.69	1.70	
51	CR	199,884	.68	1.71	
52	CR	198,858	.62	2.21	

Table M7. Mathematics Grade 3 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	200,215	.90	0.03	.31
2	MC	200,190	.89	0.04	.34
3	MC	200,154	.93	0.06	.33
4	MC	200,048	.75	0.11	.29
5	MC	200,104	.89	0.09	.32
6	MC	199,955	.54	0.16	.57
7	MC	200,030	.51	0.12	.63
8	MC	199,911	.79	0.18	.51
9	MC	200,074	.83	0.10	.48
10	MC	199,969	.84	0.15	.44
11	MC	200,026	.71	0.13	.50
12	MC	199,872	.41	0.20	.23
13	MC	199,894	.65	0.19	.44
14	MC	199,695	.52	0.29	.63
15	MC	199,706	.57	0.29	.60
16	MC	199,638	.66	0.32	.44
17	MC	199,367	.64	0.45	.51
18	MC	199,259	.61	0.51	.37
19	MC	199,022	.48	0.63	.51
20	MC	198,570	.75	0.85	.44
21	MC	200,217	.89	0.03	.43
22	MC	200,228	.96	0.02	.29
23	MC	200,149	.79	0.06	.46
24	MC	200,158	.79	0.06	.49
25	MC	200,115	.74	0.08	.51
26	MC	200,169	.66	0.05	.34
27	MC	200,129	.48	0.07	.37
28	MC	199,901	.48	0.19	.52
29	MC	199,933	.62	0.17	.59
30	MC	200,045	.86	0.12	.42
31	MC	200,019	.46	0.13	.47
32	MC	200,050	.70	0.11	.48
33	MC	200,051	.68	0.11	.48
34	MC	199,979	.84	0.15	.40
35	MC	199,874	.65	0.20	.52
36	MC	199,987	.82	0.15	.38
37	MC	199,965	.67	0.16	.48
38	MC	199,890	.52	0.19	.55
39	MC	199,740	.66	0.27	.38
40	MC	199,702	.84	0.29	.32
41	MC	198,979	.58	0.65	.54

Table M7. Mathematics Grade 3 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
42	CR	199,196	.48	0.54	
43	CR	199,543	.47	0.37	
44	CR	199,669	.61	0.30	
45	CR	199,696	.32	0.29	
46	CR	199,666	.53	0.31	
47	CR	199,821	.50	0.23	
48	CR	199,623	.36	0.33	
49	CR	199,678	.50	0.30	

Table M8. Mathematics Grade 4 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	204,736	.84	0.02	.54
2	MC	204,695	.77	0.04	.47
3	MC	204,661	.65	0.05	.41
4	MC	204,606	.75	0.08	.50
5	MC	204,668	.92	0.05	.33
6	MC	204,691	.84	0.04	.39
7	MC	204,525	.61	0.12	.43
8	MC	204,561	.40	0.10	.46
9	MC	204,611	.78	0.08	.40
10	MC	204,270	.59	0.24	.41
11	MC	204,635	.50	0.06	.57
12	MC	204,474	.76	0.14	.49
13	MC	204,571	.63	0.10	.54
14	MC	204,534	.60	0.11	.54
15	MC	204,537	.64	0.11	.53
16	MC	204,429	.79	0.17	.46
17	MC	204,472	.59	0.14	.49
18	MC	204,311	.52	0.22	.46
19	MC	203,943	.64	0.40	.56
20	MC	203,605	.56	0.57	.55
21	MC	204,742	.82	0.01	.40
22	MC	204,718	.85	0.02	.41
23	MC	204,632	.61	0.07	.66
24	MC	204,635	.70	0.06	.50
25	MC	204,641	.50	0.06	.46
26	MC	204,618	.85	0.07	.48
27	MC	204,527	.68	0.12	.56
28	MC	204,495	.44	0.13	.34
29	MC	204,550	.42	0.11	.51

Table M8. Mathematics Grade 4 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
30	MC	204,591	.74	0.09	.41
31	MC	204,546	.68	0.11	.49
32	MC	204,572	.44	0.10	.54
33	MC	204,615	.65	0.07	.62
34	MC	204,588	.65	0.09	.48
35	MC	204,486	.52	0.14	.50
36	MC	204,520	.59	0.12	.53
37	MC	204,546	.48	0.11	.44
38	MC	204,376	.44	0.19	.53
39	MC	204,157	.60	0.30	.46
40	MC	204,365	.48	0.20	.57
41	MC	204,218	.82	0.27	.46
42	MC	203,876	.79	0.44	.44
43	CR	204,610	.59	0.08	
44	CR	204,275	.51	0.24	
45	CR	204,252	.69	0.25	
46	CR	203,974	.68	0.39	
47	CR	204,241	.57	0.26	
48	CR	204,166	.60	0.29	
49	CR	204,381	.53	0.19	
50	CR	204,275	.43	0.24	
51	CR	203,913	.48	0.42	
52	CR	203,882	.56	0.43	

Table M9. Mathematics Grade 5 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	198,530	.82	0.02	.37
2	MC	198,528	.76	0.02	.40
3	MC	198,281	.84	0.15	.45
4	MC	198,493	.77	0.04	.40
5	MC	198,371	.58	0.10	.46
6	MC	198,455	.85	0.06	.48
7	MC	198,365	.41	0.10	.49
8	MC	198,362	.81	0.10	.48
9	MC	198,351	.66	0.11	.49
10	MC	198,429	.49	0.07	.50
11	MC	198,336	.51	0.12	.38
12	MC	197,966	.39	0.30	.53
13	MC	198,342	.67	0.11	.51
14	MC	198,286	.71	0.14	.48

Table M9. Mathematics Grade 5 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
15	MC	198,308	.53	0.13	.48
16	MC	198,106	.81	0.23	.47
17	MC	198,357	.75	0.11	.48
18	MC	198,171	.53	0.20	.36
19	MC	197,892	.56	0.34	.47
20	MC	197,681	.50	0.45	.63
21	MC	198,494	.62	0.04	.46
22	MC	198,477	.72	0.05	.41
23	MC	198,522	.90	0.02	.36
24	MC	198,471	.80	0.05	.48
25	MC	198,458	.72	0.06	.56
26	MC	198,377	.67	0.10	.48
27	MC	198,505	.62	0.03	.47
28	MC	198,514	.73	0.03	.24
29	MC	198,341	.63	0.12	.62
30	MC	198,209	.61	0.18	.33
31	MC	198,446	.63	0.06	.41
32	MC	198,265	.34	0.15	.35
33	MC	198,516	.85	0.03	.34
34	MC	198,200	.35	0.19	.43
35	MC	198,375	.73	0.10	.50
36	MC	198,354	.62	0.11	.49
37	MC	198,290	.64	0.14	.56
38	MC	198,212	.50	0.18	.46
39	MC	198,439	.68	0.07	.47
40	MC	198,213	.37	0.18	.32
41	MC	198,221	.75	0.18	.49
42	MC	197,626	.70	0.48	.58
43	CR	198,321	.80	0.13	
44	CR	198,307	.64	0.13	
45	CR	196,078	.38	1.25	
46	CR	198,055	.80	0.26	
47	CR	197,972	.51	0.30	
48	CR	198,052	.48	0.26	
49	CR	197,406	.28	0.59	
50	CR	197,328	.26	0.63	
51	CR	197,331	.20	0.62	
52	CR	197,159	.38	0.71	

Table M10. Mathematics Grade 6 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	199,057	.80	0.04	.55
2	MC	199,083	.73	0.02	.51
3	MC	198,843	.59	0.14	.56
4	MC	199,060	.79	0.04	.49
5	MC	198,910	.84	0.11	.49
6	MC	198,738	.51	0.20	.56
7	MC	198,917	.58	0.11	.31
8	MC	199,080	.84	0.03	.41
9	MC	198,724	.53	0.20	.34
10	MC	198,846	.56	0.14	.47
11	MC	198,943	.62	0.09	.56
12	MC	199,030	.52	0.05	.29
13	MC	198,662	.56	0.24	.45
14	MC	198,979	.72	0.08	.53
15	MC	198,805	.38	0.16	.36
16	MC	198,894	.57	0.12	.41
17	MC	198,809	.67	0.16	.43
18	MC	198,827	.40	0.15	.34
19	MC	198,815	.75	0.16	.50
20	MC	198,831	.75	0.15	.47
21	MC	198,491	.60	0.32	.42
22	MC	198,808	.91	0.16	.33
23	MC	198,702	.56	0.21	.36
24	MC	198,117	.80	0.51	.56
25	MC	199,070	.87	0.03	.40
26	MC	199,101	.87	0.01	.36
27	MC	198,906	.75	0.11	.43
28	MC	198,895	.55	0.12	.64
29	MC	199,056	.72	0.04	.52
30	MC	198,903	.65	0.11	.48
31	MC	199,071	.37	0.03	.53
32	MC	198,555	.38	0.29	.36
33	MC	198,810	.56	0.16	.41
34	MC	198,934	.59	0.10	.46
35	MC	198,811	.55	0.16	.41
36	MC	199,036	.71	0.05	.50
37	MC	198,997	.80	0.07	.46
38	MC	198,947	.50	0.09	.55
39	MC	198,943	.54	0.09	.47
40	MC	198,879	.62	0.13	.55
41	MC	198,844	.70	0.14	.32

Table M10. Mathematics Grade 6 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
42	MC	198,887	.46	0.12	.44
43	MC	198,766	.42	0.18	.47
44	MC	198,929	.67	0.10	.44
45	MC	198,457	.43	0.34	.35
46	MC	198,837	.52	0.15	.57
47	MC	198,774	.79	0.18	.44
48	MC	197,678	.70	0.73	.36
49	CR	197,812	.60	0.66	
50	CR	198,644	.48	0.24	
51	CR	198,259	.29	0.44	
52	CR	198,008	.45	0.56	
53	CR	197,117	.65	1.01	
54	CR	198,198	.65	0.47	
55	CR	197,070	.20	1.03	
56	CR	198,246	.38	0.44	
57	CR	196,630	.19	1.26	
58	CR	197,358	.71	0.89	

Table M11. Mathematics Grade 7 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	193,292	.84	0.03	.29
2	MC	193,235	.71	0.06	.44
3	MC	192,786	.73	0.29	.46
4	MC	193,133	.54	0.11	.47
5	MC	193,151	.44	0.10	.26
6	MC	193,111	.39	0.12	.27
7	MC	192,797	.37	0.25	.44
8	MC	193,020	.42	0.17	.40
9	MC	193,184	.70	0.08	.45
10	MC	193,115	.55	0.12	.36
11	MC	193,109	.28	0.12	.41
12	MC	193,077	.67	0.14	.52
13	MC	192,964	.38	0.20	.55
14	MC	193,137	.39	0.11	.41
15	MC	192,958	.55	0.20	.50
16	MC	193,081	.70	0.14	.35
17	MC	193,050	.62	0.15	.56
18	MC	193,012	.68	0.17	.40
19	MC	193,065	.74	0.14	.44
20	MC	192,708	.33	0.33	.33

Table M11. Mathematics Grade 7 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
21	MC	192,810	.49	0.28	.42
22	MC	192,729	.45	0.32	.23
23	MC	192,555	.56	0.41	.39
24	MC	192,255	.51	0.56	.50
25	MC	193,010	.67	0.17	.45
26	MC	193,269	.80	0.04	.49
27	MC	192,767	.55	0.30	.44
28	MC	193,252	.74	0.05	.38
29	MC	192,998	.58	0.18	.52
30	MC	193,217	.82	0.07	.36
31	MC	193,257	.40	0.04	.57
32	MC	193,132	.36	0.11	.28
33	MC	192,973	.39	0.19	.30
34	MC	193,198	.22	0.07	.43
35	MC	193,195	.46	0.08	.53
36	MC	193,085	.58	0.13	.53
37	MC	193,164	.52	0.09	.48
38	MC	193,172	.79	0.09	.47
39	MC	193,083	.51	0.13	.54
40	MC	193,013	.54	0.17	.33
41	MC	192,928	.53	0.21	.54
42	MC	193,180	.59	0.08	.46
43	MC	193,156	.59	0.10	.43
44	MC	193,015	.49	0.17	.41
45	MC	193,082	.75	0.13	.50
46	MC	193,058	.72	0.15	.51
47	MC	192,972	.51	0.19	.59
48	MC	192,659	.80	0.35	.43
49	CR	192,489	.66	0.44	
50	CR	191,924	.56	0.73	
51	CR	189,977	.29	1.74	
52	CR	191,701	.50	0.85	
53	CR	190,954	.50	1.24	
54	CR	190,916	.50	1.26	
55	CR	192,196	.28	0.59	
56	CR	191,399	.46	1.01	
57	CR	190,137	.28	1.66	
58	CR	191,533	.55	0.94	

Table M12. Mathematics Grade 8 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	159,168	.74	0.09	.40
2	MC	159,216	.71	0.06	.42
3	MC	159,187	.52	0.07	.35
4	MC	159,127	.50	0.11	.41
5	MC	159,138	.70	0.10	.43
6	MC	159,198	.57	0.07	.42
7	MC	159,013	.33	0.18	.40
8	MC	159,197	.58	0.07	.40
9	MC	158,761	.45	0.34	.43
10	MC	159,203	.80	0.06	.30
11	MC	159,049	.32	0.16	.43
12	MC	159,076	.44	0.14	.43
13	MC	159,163	.63	0.09	.41
14	MC	159,180	.67	0.08	.51
15	MC	159,024	.35	0.18	.50
16	MC	159,154	.80	0.09	.44
17	MC	159,066	.51	0.15	.39
18	MC	159,169	.42	0.08	.22
19	MC	158,923	.46	0.24	.39
20	MC	159,129	.68	0.11	.40
21	MC	159,104	.62	0.13	.27
22	MC	159,112	.62	0.12	.46
23	MC	159,016	.43	0.18	.31
24	MC	158,856	.63	0.28	.39
25	MC	159,289	.89	0.01	.31
26	MC	158,962	.53	0.21	.45
27	MC	159,240	.63	0.04	.38
28	MC	158,972	.34	0.21	.47
29	MC	159,141	.42	0.10	.15
30	MC	159,140	.57	0.10	.49
31	MC	159,138	.49	0.10	.29
32	MC	159,165	.46	0.09	.41
33	MC	159,103	.63	0.13	.44
34	MC	159,201	.69	0.06	.35
35	MC	159,180	.65	0.08	.43
36	MC	158,921	.37	0.24	.40
37	MC	159,196	.69	0.07	.35
38	MC	159,109	.44	0.12	.30
39	MC	159,208	.68	0.06	.41
40	MC	159,028	.36	0.17	.33
41	MC	159,099	.56	0.13	.43

Table M12. Mathematics Grade 8 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
42	MC	159,165	.75	0.09	.45
43	MC	159,162	.53	0.09	.43
44	MC	159,217	.65	0.05	.39
45	MC	159,134	.52	0.11	.45
46	MC	159,094	.52	0.13	.45
47	MC	159,166	.53	0.09	.52
48	MC	159,068	.65	0.15	.32
49	CR	158,253	.60	0.66	
50	CR	155,870	.32	2.16	
51	CR	154,424	.28	3.06	
52	CR	153,936	.28	3.37	
53	CR	155,684	.20	2.27	
54	CR	154,696	.33	2.89	
55	CR	156,419	.38	1.81	
56	CR	156,308	.45	1.88	
57	CR	154,684	.33	2.90	
58	CR	155,755	.51	2.23	

Appendix N: Items Flagged for DIF

These tables support the DIF information in Section 5, “Operational Test Data Collection and Classical Analysis.” They include item numbers, focal group, and directions of DIF and DIF statistics. Tables N1–N4 show items flagged by the SMD, or Mantel-Haenszel methods. Note that positive values of SMD and Delta in Tables N1–N4 indicate DIF in favor of a focal group, and negative values of SMD and Delta indicate DIF against a focal group. External linking and field test items (i.e., those not contributing to students’ scores) have been omitted.

Table N1. ELA MC Item Classical DIF Flags

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
4	3	Female	Against	1.65	1,863.1	-1.17
4	8	Female	Against	1.86	3,242.1	-1.45
4	13	Black	Against	1.66	1,341.4	-1.19
4	13	Hispanic	Against	1.70	1,844.7	-1.24
4	21	Female	Against	1.59	2,298.4	-1.09
5	6	Hispanic	Against	1.86	2,244.8	-1.46
5	6	High Needs	Against	1.65	1,986.8	-1.18
5	10	Female	Against	1.75	3,080.5	-1.31
5	13	Female	Against	1.55	1,724.1	-1.03
5	22	Female	Against	1.77	1,940.3	-1.34
5	39	Female	Against	1.86	3,663.6	-1.46
6	10	Hispanic	Against	1.63	455.9	-1.15
6	10	High Needs	Against	1.66	530.3	-1.19
6	17	Black	Against	1.61	655.9	-1.11
6	17	Hispanic	Against	1.59	711.1	-1.09
6	33	Hispanic	Against	1.99	2,503.1	-1.61
6	33	High Needs	Against	1.70	1,963.6	-1.25
7	14	Female	Against	1.66	2,297.6	-1.20
7	29	Female	Against	1.69	2,490.3	-1.23
7	31	Black	Against	1.56	801.6	-1.04
7	35	Hispanic	Against	1.56	1,096.2	-1.04
8	9	Female	Against	1.64	2,350.7	-1.16
8	16	Female	Against	1.53	815.5	-1.00
8	22	Black	Against	1.60	1,119.8	-1.10
8	22	Hispanic	Against	1.66	1,508.3	-1.19
8	25	Black	Against	1.70	1,645.5	-1.24
8	36	Hispanic	Against	1.66	1,685.1	-1.19

Table N2. ELA CR Item Classical DIF Flags

Grade	Item	Subgroup	DIF	SMD	Effect Size
4	39	Hispanic	In Favor	0.1155	0.171
6	47	Black	In Favor	0.1324	0.194
6	48	Black	In Favor	0.1358	0.192
7	45	Female	In Favor	0.1836	0.192
8	48	Black	In Favor	0.1109	0.179
8	48	Hispanic	In Favor	0.1120	0.181
8	51	Female	In Favor	0.1156	0.175
8	52	Female	In Favor	0.2334	0.230

Table N3. Mathematics MC Item Classical DIF Flags

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
3	15	Hispanic	Against	1.73	1,525.3	-1.28
4	2	Black	Against	1.56	794.0	-1.05
4	13	Female	Against	1.82	2,919.1	-1.40
4	18	Black	Against	1.60	1,081.0	-1.10
5	12	Female	Against	1.55	1,534.6	-1.03
5	28	Black	In Favor	0.65	827.7	1.03
6	4	Female	Against	1.86	2,195.5	-1.46
6	4	Black	Against	1.83	1,206.4	-1.42
6	4	Hispanic	Against	1.91	1,617.5	-1.52
6	4	High Needs	Against	1.61	985.7	-1.12
6	31	Black	In Favor	0.62	832.4	1.11
7	1	Female	Against	1.63	1,374.1	-1.15
7	15	Female	Against	1.57	1,821.4	-1.07
7	34	Female	Against	1.53	988.1	-1.00
7	43	Hispanic	Against	1.64	1,481.0	-1.16
7	43	High Needs	Against	1.78	2,676.9	-1.36
8	14	Female	Against	2.06	3,136.6	-1.70
8	28	Black	In Favor	0.64	671.8	1.04
8	34	Female	Against	1.61	1,680.9	-1.12

Table N4. Mathematics CR Item Classical DIF Flags

Grade	Item	Subgroup	DIF	SMD	Effect Size
3	49	Female	Against	-0.2424	-0.201
3	49	Black	Against	-0.2734	-0.240

Appendix O: Item Response Theory Statistics

External linking and field test items (i.e., those not contributing to students' scores) have been omitted.

Table O1. ELA Grade 3 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	595.22	7	157.21	168.86	Y
2	3PL	283.67	7	73.94	168.78	Y
3	3PL	70.90	7	17.08	168.74	Y
4	3PL	174.24	7	44.70	168.74	Y
5	3PL	134.91	7	34.19	168.74	Y
6	3PL	94.56	7	23.40	168.65	Y
7	3PL	82.52	7	20.18	168.71	Y
8	3PL	109.15	7	27.30	168.61	Y
9	3PL	180.99	7	46.50	168.71	Y
10	3PL	131.22	7	33.20	168.61	Y
11	3PL	33.70	7	7.14	168.57	Y
12	3PL	165.77	7	42.43	168.50	Y
13	3PL	114.42	7	28.71	168.20	Y
14	3PL	144.84	7	36.84	168.12	Y
15	3PL	117.28	7	29.47	168.05	Y
16	3PL	116.90	7	29.37	167.91	Y
17	3PL	208.92	7	53.96	167.78	Y
18	3PL	105.71	7	26.38	167.46	Y
19	3PL	230.72	7	59.79	162.43	Y
20	3PL	151.49	7	38.62	161.55	Y
21	3PL	148.60	7	37.85	160.93	Y
22	3PL	146.78	7	37.36	160.50	Y
23	3PL	99.88	7	24.82	168.78	Y
24	3PL	113.90	7	28.57	168.76	Y
25	3PL	213.20	7	55.11	168.69	Y
26	3PL	79.38	7	19.34	168.58	Y
27	3PL	110.52	7	27.67	168.69	Y
28	3PL	90.51	7	22.32	168.57	Y
29	3PL	95.87	7	23.75	168.28	Y

Table O1. ELA Grade 3 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
30	2PPC	539.36	16	92.52	167.69	Y
31	2PPC	513.83	16	88.00	165.79	Y
32	2PPC	253.86	16	42.05	163.89	Y
33	2PPC	1125.38	34	132.35	161.90	Y
34	2PPC	598.08	16	102.90	166.61	Y
35	2PPC	279.71	16	46.62	162.45	Y
36	2PPC	503.54	16	86.19	165.50	Y
37	2PPC	399.11	16	67.73	163.18	Y
38	2PPC	507.36	16	86.86	162.00	Y

Table O2. ELA Grade 4 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	83.25	7	20.38	171.93	Y
2	3PL	108.57	7	27.15	171.94	Y
3	3PL	80.50	7	19.64	171.92	Y
4	3PL	69.49	7	16.70	171.77	Y
5	3PL	82.55	7	20.19	171.87	Y
6	3PL	112.51	7	28.20	171.85	Y
7	3PL	115.55	7	29.01	171.86	Y
8	3PL	217.94	7	56.38	171.86	Y
9	3PL	100.37	7	24.95	171.85	Y
10	3PL	139.62	7	35.44	171.81	Y
11	3PL	169.35	7	43.39	171.79	Y
12	3PL	211.95	7	54.77	171.81	Y
13	3PL	225.15	7	58.30	171.71	Y
14	3PL	232.25	7	60.20	171.67	Y
15	3PL	193.81	7	49.93	171.62	Y
16	3PL	143.44	7	36.47	171.64	Y
17	3PL	209.32	7	54.07	171.59	Y
18	3PL	209.46	7	54.11	171.50	Y
19	3PL	90.38	7	22.28	169.75	Y
20	3PL	66.87	7	16.00	169.61	Y
21	3PL	89.49	7	22.05	169.28	Y
22	3PL	159.62	7	40.79	169.06	Y
23	3PL	74.10	7	17.93	168.85	Y
24	3PL	45.35	7	10.25	168.70	Y
25	3PL	68.29	7	16.38	171.79	Y
26	3PL	83.56	7	20.46	171.79	Y
27	3PL	74.33	7	17.99	171.74	Y
28	3PL	164.00	7	41.96	171.69	Y

Table O2. ELA Grade 4 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
29	3PL	22.76	7	4.21	171.75	Y
30	3PL	110.50	7	27.66	171.75	Y
31	3PL	102.91	7	25.63	171.55	Y
32	2PPC	397.29	16	67.40	170.68	Y
33	2PPC	622.55	16	107.22	170.77	Y
34	2PPC	536.97	16	92.09	169.51	Y
35	2PPC	988.80	34	115.79	167.55	Y
36	2PPC	403.08	16	68.43	171.21	Y
37	2PPC	516.28	16	88.44	170.75	Y
38	2PPC	530.75	16	91.00	171.02	Y
39	2PPC	420.37	16	71.48	169.56	Y
40	2PPC	344.16	16	58.01	168.79	Y
41	2PPC	975.00	34	114.11	168.46	Y

Table O3. ELA Grade 5 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	86.60	7	21.27	169.26	Y
2	3PL	41.43	7	9.20	169.15	Y
3	3PL	90.35	7	22.28	169.21	Y
4	3PL	122.01	7	30.74	169.05	Y
5	3PL	67.33	7	16.12	169.17	Y
6	3PL	150.69	7	38.40	169.14	Y
7	3PL	136.93	7	34.72	169.15	Y
8	3PL	102.85	7	25.62	169.10	Y
9	3PL	325.33	7	85.08	169.15	Y
10	3PL	111.72	7	27.99	169.09	Y
11	3PL	58.50	7	13.76	168.99	Y
12	3PL	114.24	7	28.66	169.10	Y
13	3PL	99.18	7	24.64	169.04	Y
14	3PL	115.61	7	29.03	169.09	Y
15	3PL	53.25	7	12.36	169.02	Y
16	3PL	92.95	7	22.97	169.09	Y
17	3PL	69.03	7	16.58	168.98	Y
18	3PL	166.01	7	42.50	168.98	Y
19	3PL	161.86	7	41.39	168.81	Y
20	3PL	189.46	7	48.76	168.92	Y
21	3PL	134.18	7	33.99	168.87	Y
22	3PL	95.81	7	23.74	168.66	Y
23	3PL	112.70	7	28.25	168.60	Y
24	3PL	79.78	7	19.45	168.50	Y

Table O3. ELA Grade 5 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
25	3PL	221.57	7	57.35	168.37	Y
26	3PL	240.98	7	62.53	168.37	Y
27	3PL	173.58	7	44.52	168.17	Y
28	3PL	198.62	7	51.21	168.11	Y
29	3PL	157.74	7	40.29	167.19	Y
30	3PL	125.20	7	31.59	167.17	Y
31	3PL	119.54	7	30.08	166.96	Y
32	3PL	138.34	7	35.10	166.78	Y
33	3PL	83.86	7	20.54	166.59	Y
34	3PL	89.08	7	21.94	166.31	Y
35	3PL	205.78	7	53.13	166.07	Y
36	3PL	55.26	7	12.90	169.08	Y
37	3PL	73.12	7	17.67	169.07	Y
38	3PL	99.53	7	24.73	169.03	Y
39	3PL	64.02	7	15.24	169.04	Y
40	3PL	71.65	7	17.28	169.08	Y
41	3PL	59.11	7	13.93	169.06	Y
42	3PL	113.03	7	28.34	168.98	Y
43	2PPC	529.95	16	90.85	168.67	Y
44	2PPC	322.43	16	54.17	168.36	Y
45	2PPC	482.21	16	82.42	167.97	Y
46	2PPC	1011.58	34	118.55	167.18	Y
47	2PPC	227.33	16	37.36	168.77	Y
48	2PPC	447.83	16	76.34	167.34	Y
49	2PPC	318.62	16	53.50	168.22	Y
50	2PPC	963.68	16	167.53	165.99	N
51	2PPC	405.34	16	68.83	166.64	Y
52	2PPC	769.49	34	89.19	165.73	Y

Table O4. ELA Grade 6 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	105.33	7	26.28	169.69	Y
2	3PL	97.81	7	24.27	169.53	Y
3	3PL	149.06	7	37.97	169.67	Y
4	3PL	110.07	7	27.55	169.62	Y
5	3PL	62.11	7	14.73	169.59	Y
6	3PL	77.88	7	18.94	169.59	Y
7	3PL	86.47	7	21.24	169.62	Y
8	3PL	132.80	7	33.62	169.58	Y
9	3PL	92.21	7	22.77	169.48	Y
10	3PL	129.18	7	32.65	169.62	Y
11	3PL	215.13	7	55.62	169.56	Y
12	3PL	27.94	7	5.60	169.56	Y
13	3PL	73.02	7	17.64	169.40	Y
14	3PL	47.90	7	10.93	169.54	Y
15	3PL	110.54	7	27.67	169.26	Y
16	3PL	100.07	7	24.87	169.40	Y
17	3PL	110.46	7	27.65	169.33	Y
18	3PL	58.06	7	13.65	169.21	Y
19	3PL	88.48	7	21.78	169.25	Y
20	3PL	62.29	7	14.78	169.22	Y
21	3PL	426.86	7	112.21	168.72	Y
22	3PL	166.50	7	42.63	168.78	Y
23	3PL	111.35	7	27.89	168.67	Y
24	3PL	126.74	7	32.00	168.63	Y
25	3PL	160.36	7	40.99	168.50	Y
26	3PL	116.49	7	29.26	168.32	Y
27	3PL	89.60	7	22.08	168.24	Y
28	3PL	122.88	7	30.97	167.66	Y
29	3PL	192.25	7	49.51	167.45	Y
30	3PL	119.33	7	30.02	167.23	Y
31	3PL	92.85	7	22.94	167.05	Y
32	3PL	100.90	7	25.10	166.72	Y
33	3PL	241.35	7	62.63	166.85	Y
34	3PL	177.29	7	45.51	166.60	Y
35	3PL	109.81	7	27.48	166.47	Y
36	3PL	82.15	7	20.08	169.51	Y
37	3PL	120.41	7	30.31	169.49	Y
38	3PL	51.67	7	11.94	169.44	Y
39	3PL	47.38	7	10.79	169.45	Y
40	3PL	127.91	7	32.31	169.48	Y
41	3PL	130.64	7	33.04	169.44	Y

Table O4. ELA Grade 6 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
42	3PL	54.80	7	12.78	169.37	Y
43	2PPC	286.31	16	47.79	169.04	Y
44	2PPC	456.26	16	77.83	167.59	Y
45	2PPC	307.31	16	51.50	168.05	Y
46	2PPC	670.03	34	77.13	166.27	Y
47	2PPC	250.15	16	41.39	168.65	Y
48	2PPC	225.71	16	37.07	168.50	Y
49	2PPC	326.99	16	54.98	168.26	Y
50	2PPC	266.38	16	44.26	167.65	Y
51	2PPC	260.90	16	43.29	166.75	Y
52	2PPC	750.01	34	86.83	165.64	Y

Table O5. ELA Grade 7 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	71.82	7	17.32	167.75	Y
2	3PL	99.12	7	24.62	167.57	Y
3	3PL	51.14	7	11.80	167.54	Y
4	3PL	72.29	7	17.45	167.68	Y
5	3PL	38.28	7	8.36	167.61	Y
6	3PL	113.04	7	28.34	167.74	Y
7	3PL	77.00	7	18.71	167.51	Y
8	3PL	121.17	7	30.51	167.61	Y
9	3PL	148.06	7	37.70	167.53	Y
10	3PL	177.91	7	45.68	167.58	Y
11	3PL	47.88	7	10.93	167.61	Y
12	3PL	75.00	7	18.17	167.52	Y
13	3PL	82.98	7	20.31	167.32	Y
14	3PL	66.93	7	16.02	167.58	Y
15	3PL	27.09	7	5.37	167.50	Y
16	3PL	133.33	7	33.76	167.49	Y
17	3PL	165.38	7	42.33	167.46	Y
18	3PL	68.86	7	16.53	167.47	Y
19	3PL	146.92	7	37.40	167.46	Y
20	3PL	188.78	7	48.58	166.77	Y
21	3PL	390.89	7	102.60	166.62	Y
22	3PL	41.12	7	9.12	166.49	Y
23	3PL	270.63	7	70.46	166.49	Y
24	3PL	134.02	7	33.95	166.25	Y
25	3PL	259.62	7	67.52	166.14	Y
26	3PL	437.35	7	115.02	166.00	Y

Table O5. ELA Grade 7 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
27	3PL	201.11	7	51.88	165.84	Y
28	3PL	103.17	7	25.70	165.09	Y
29	3PL	155.34	7	39.65	164.91	Y
30	3PL	192.37	7	49.54	164.71	Y
31	3PL	97.33	7	24.14	164.52	Y
32	3PL	149.83	7	38.17	164.20	Y
33	3PL	52.40	7	12.13	164.07	Y
34	3PL	85.03	7	20.86	163.91	Y
35	3PL	67.25	7	16.10	167.51	Y
36	3PL	92.26	7	22.79	167.49	Y
37	3PL	91.11	7	22.48	167.40	Y
38	3PL	99.38	7	24.69	167.43	Y
39	3PL	55.83	7	13.05	167.45	Y
40	3PL	39.05	7	8.57	167.40	Y
41	3PL	92.65	7	22.89	167.36	Y
42	2PPC	370.36	16	62.64	166.77	Y
43	2PPC	465.71	16	79.50	166.30	Y
44	2PPC	514.87	16	88.19	165.44	Y
45	2PPC	791.19	34	91.82	165.02	Y
46	2PPC	325.32	16	54.68	166.45	Y
47	2PPC	229.88	16	37.81	165.97	Y
48	2PPC	581.01	16	99.88	165.17	Y
49	2PPC	396.34	16	67.23	164.93	Y
50	2PPC	592.82	16	101.97	162.78	Y
51	2PPC	963.70	34	112.74	162.75	Y

Table O6. ELA Grade 8 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	608.31	7	160.71	173.00	Y
2	3PL	144.56	7	36.77	172.89	Y
3	3PL	98.55	7	24.47	172.94	Y
4	3PL	73.73	7	17.83	172.91	Y
5	3PL	208.39	7	53.82	172.75	Y
6	3PL	77.88	7	18.94	172.81	Y
7	3PL	138.78	7	35.22	172.86	Y
8	3PL	54.91	7	12.81	172.73	Y
9	3PL	125.87	7	31.77	172.72	Y
10	3PL	118.78	7	29.87	172.67	Y
11	3PL	94.48	7	23.38	172.85	Y
12	3PL	107.95	7	26.98	172.80	Y

Table O6. ELA Grade 8 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
13	3PL	126.88	7	32.04	172.79	Y
14	3PL	103.80	7	25.87	172.78	Y
15	3PL	106.30	7	26.54	172.82	Y
16	3PL	129.40	7	32.71	172.92	Y
17	3PL	72.77	7	17.58	172.75	Y
18	3PL	119.81	7	30.15	172.82	Y
19	3PL	162.57	7	41.58	172.82	Y
20	3PL	90.01	7	22.19	172.81	Y
21	3PL	86.02	7	21.12	172.74	Y
22	3PL	138.41	7	35.12	172.36	Y
23	3PL	93.47	7	23.11	172.29	Y
24	3PL	145.99	7	37.15	172.14	Y
25	3PL	74.27	7	17.98	172.13	Y
26	3PL	190.33	7	49.00	172.00	Y
27	3PL	232.87	7	60.37	171.94	Y
28	3PL	126.08	7	31.83	171.88	Y
29	3PL	130.11	7	32.90	171.56	Y
30	3PL	191.16	7	49.22	171.45	Y
31	3PL	88.19	7	21.70	171.14	Y
32	3PL	54.07	7	12.58	171.02	Y
33	3PL	101.51	7	25.26	170.96	Y
34	3PL	137.17	7	34.79	170.81	Y
35	3PL	170.56	7	43.71	170.74	Y
36	3PL	75.47	7	18.30	172.55	Y
37	3PL	50.01	7	11.49	172.54	Y
38	3PL	124.63	7	31.44	172.56	Y
39	3PL	98.56	7	24.47	172.49	Y
40	3PL	85.07	7	20.86	172.53	Y
41	3PL	118.41	7	29.78	172.47	Y
42	3PL	89.58	7	22.07	172.41	Y
43	2PPC	334.53	16	56.31	171.56	Y
44	2PPC	288.97	16	48.26	170.31	Y
45	2PPC	348.80	16	58.83	170.78	Y
46	2PPC	784.48	34	91.01	169.14	Y
47	2PPC	164.41	16	26.24	171.65	Y
48	2PPC	238.97	16	39.42	171.52	Y
49	2PPC	478.23	16	81.71	171.26	Y
50	2PPC	404.52	16	68.68	169.32	Y
51	2PPC	528.84	16	90.66	169.26	Y
52	2PPC	1071.74	34	125.84	168.53	Y

Table O7. Mathematics Grade 3 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	153.91	7	39.26	176.00	Y
2	3PL	105.77	7	26.40	175.98	Y
3	3PL	61.03	7	14.44	175.95	Y
4	3PL	235.63	7	61.10	175.87	Y
5	3PL	117.41	7	29.51	175.88	Y
6	3PL	191.79	7	49.39	175.73	Y
7	3PL	212.93	7	55.04	175.84	Y
8	3PL	220.59	7	57.08	175.75	Y
9	3PL	108.48	7	27.12	175.85	Y
10	3PL	299.25	7	78.11	175.75	Y
11	3PL	333.55	7	87.27	175.81	Y
12	3PL	51.16	7	11.80	175.68	Y
13	3PL	156.26	7	39.89	175.73	Y
14	3PL	204.26	7	52.72	175.57	Y
15	3PL	173.12	7	44.40	175.52	Y
16	3PL	179.83	7	46.19	175.50	Y
17	3PL	103.00	7	25.66	175.27	Y
18	3PL	150.41	7	38.33	175.14	Y
19	3PL	193.31	7	49.79	174.96	Y
20	3PL	87.04	7	21.39	174.54	Y
21	3PL	127.32	7	32.16	175.98	Y
22	3PL	87.58	7	21.54	176.02	Y
23	3PL	91.83	7	22.67	175.96	Y
24	3PL	83.41	7	20.42	175.95	Y
25	3PL	171.68	7	44.01	175.91	Y
26	3PL	140.75	7	35.75	175.95	Y
27	3PL	86.06	7	21.13	175.93	Y
28	3PL	225.13	7	58.30	175.75	Y
29	3PL	195.38	7	50.35	175.74	Y
30	3PL	119.26	7	30.00	175.83	Y
31	3PL	200.55	7	51.73	175.84	Y
32	3PL	91.80	7	22.66	175.86	Y
33	3PL	149.49	7	38.08	175.86	Y
34	3PL	195.70	7	50.43	175.78	Y
35	3PL	164.54	7	42.11	175.69	Y
36	3PL	83.59	7	20.47	175.82	Y
37	3PL	97.01	7	24.05	175.79	Y
38	3PL	188.88	7	48.61	175.69	Y
39	3PL	154.27	7	39.36	175.60	Y
40	3PL	106.94	7	26.71	175.55	Y
41	3PL	148.35	7	37.78	174.89	Y

Table O7. Mathematics Grade 3 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
42	2PPC	1608.55	16	281.53	175.15	N
43	2PPC	837.04	16	145.14	175.44	Y
44	2PPC	3158.09	16	555.45	175.57	N
45	2PPC	257.92	16	42.77	175.56	Y
46	2PPC	543.12	16	93.18	175.51	Y
47	2PPC	2893.60	25	405.68	175.62	N
48	2PPC	736.45	25	100.61	175.48	Y
49	2PPC	415.77	25	55.26	175.52	Y

Table O8. Mathematics Grade 4 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	94.42	7	23.36	179.97	Y
2	3PL	52.44	7	12.14	179.93	Y
3	3PL	87.49	7	21.51	179.90	Y
4	3PL	54.29	7	12.64	179.87	Y
5	3PL	75.29	7	18.25	179.91	Y
6	3PL	112.05	7	28.07	179.95	Y
7	3PL	69.20	7	16.62	179.81	Y
8	3PL	161.94	7	41.41	179.85	Y
9	3PL	117.44	7	29.52	179.89	Y
10	3PL	76.02	7	18.45	179.57	Y
11	3PL	153.61	7	39.18	179.89	Y
12	3PL	65.98	7	15.76	179.75	Y
13	3PL	129.50	7	32.74	179.84	Y
14	3PL	109.61	7	27.42	179.81	Y
15	3PL	74.07	7	17.93	179.81	Y
16	3PL	60.53	7	14.31	179.72	Y
17	3PL	103.81	7	25.87	179.74	Y
18	3PL	89.87	7	22.15	179.66	Y
19	3PL	96.14	7	23.82	179.33	Y
20	3PL	92.61	7	22.88	178.99	Y
21	3PL	129.93	7	32.85	179.99	Y
22	3PL	133.05	7	33.69	179.97	Y
23	3PL	160.22	7	40.95	179.87	Y
24	3PL	132.79	7	33.62	179.90	Y
25	3PL	105.33	7	26.28	179.93	Y
26	3PL	191.04	7	49.19	179.88	Y
27	3PL	86.64	7	21.29	179.83	Y
28	3PL	112.78	7	28.27	179.78	Y
29	3PL	181.96	7	46.76	179.81	Y

Table O8. Mathematics Grade 4 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
30	3PL	131.46	7	33.26	179.86	Y
31	3PL	62.54	7	14.84	179.78	Y
32	3PL	117.28	7	29.47	179.82	Y
33	3PL	147.93	7	37.67	179.85	Y
34	3PL	173.93	7	44.61	179.85	Y
35	3PL	110.31	7	27.61	179.79	Y
36	3PL	128.90	7	32.58	179.79	Y
37	3PL	100.21	7	24.91	179.81	Y
38	3PL	218.73	7	56.59	179.67	Y
39	3PL	118.00	7	29.67	179.51	Y
40	3PL	189.47	7	48.77	179.70	Y
41	3PL	67.57	7	16.19	179.55	Y
42	3PL	85.64	7	21.02	179.27	Y
43	2PPC	295.42	16	49.39	179.85	Y
44	2PPC	393.02	16	66.65	179.57	Y
45	2PPC	318.83	16	53.53	179.53	Y
46	2PPC	1700.81	16	297.84	179.25	N
47	2PPC	460.31	16	78.54	179.46	Y
48	2PPC	406.17	16	68.97	179.46	Y
49	2PPC	931.94	25	128.26	179.63	Y
50	2PPC	2093.71	25	292.56	179.54	N
51	2PPC	1989.00	25	277.75	179.30	N
52	2PPC	1141.45	25	157.89	179.27	Y

Table O9. Mathematics Grade 5 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	92.38	7	22.82	174.51	Y
2	3PL	49.56	7	11.37	174.50	Y
3	3PL	308.57	7	80.60	174.28	Y
4	3PL	64.89	7	15.47	174.48	Y
5	3PL	113.18	7	28.38	174.34	Y
6	3PL	165.35	7	42.32	174.41	Y
7	3PL	102.03	7	25.40	174.35	Y
8	3PL	263.70	7	68.60	174.38	Y
9	3PL	101.00	7	25.12	174.34	Y
10	3PL	98.72	7	24.51	174.42	Y
11	3PL	37.45	7	8.14	174.34	Y
12	3PL	174.70	7	44.82	174.08	Y
13	3PL	76.58	7	18.60	174.37	Y
14	3PL	229.78	7	59.54	174.30	Y

Table O9. Mathematics Grade 5 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
15	3PL	248.32	7	64.50	174.31	Y
16	3PL	286.24	7	74.63	174.15	Y
17	3PL	110.07	7	27.55	174.37	Y
18	3PL	128.50	7	32.47	174.21	Y
19	3PL	162.73	7	41.62	173.94	Y
20	3PL	156.00	7	39.82	173.81	Y
21	3PL	74.32	7	17.99	174.49	Y
22	3PL	64.32	7	15.32	174.46	Y
23	3PL	96.33	7	23.87	174.49	Y
24	3PL	114.02	7	28.60	174.43	Y
25	3PL	134.64	7	34.11	174.46	Y
26	3PL	103.74	7	25.85	174.37	Y
27	3PL	73.98	7	17.90	174.50	Y
28	3PL	95.27	7	23.59	174.49	Y
29	3PL	160.58	7	41.04	174.34	Y
30	3PL	77.48	7	18.84	174.23	Y
31	3PL	42.30	7	9.43	174.46	Y
32	3PL	117.40	7	29.51	174.29	Y
33	3PL	157.51	7	40.23	174.50	Y
34	3PL	157.51	7	40.23	174.20	Y
35	3PL	129.48	7	32.74	174.38	Y
36	3PL	65.92	7	15.75	174.37	Y
37	3PL	129.03	7	32.62	174.29	Y
38	3PL	119.85	7	30.16	174.21	Y
39	3PL	120.43	7	30.31	174.44	Y
40	3PL	100.04	7	24.86	174.23	Y
41	3PL	151.90	7	38.72	174.23	Y
42	3PL	120.70	7	30.39	173.71	Y
43	2PPC	337.46	16	56.83	174.36	Y
44	2PPC	969.51	16	168.56	174.31	Y
45	2PPC	1144.01	16	199.41	172.45	N
46	2PPC	1009.04	16	175.55	174.14	N
47	2PPC	253.80	16	42.04	174.01	Y
48	2PPC	596.29	16	102.58	174.06	Y
49	2PPC	953.05	25	131.25	173.48	Y
50	2PPC	620.35	25	84.20	173.39	Y
51	2PPC	910.80	25	125.27	173.42	Y
52	2PPC	1018.82	25	140.55	173.20	Y

Table O10. Mathematics Grade 6 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	195.12	7	50.28	174.99	Y
2	3PL	123.47	7	31.13	175.06	Y
3	3PL	158.70	7	40.54	174.83	Y
4	3PL	93.82	7	23.20	175.02	Y
5	3PL	225.13	7	58.30	174.87	Y
6	3PL	74.63	7	18.07	174.75	Y
7	3PL	103.29	7	25.73	174.89	Y
8	3PL	96.51	7	23.92	175.04	Y
9	3PL	95.95	7	23.77	174.72	Y
10	3PL	139.98	7	35.54	174.83	Y
11	3PL	145.65	7	37.06	174.94	Y
12	3PL	52.54	7	12.17	174.97	Y
13	3PL	98.95	7	24.58	174.67	Y
14	3PL	112.92	7	28.31	174.95	Y
15	3PL	73.96	7	17.89	174.84	Y
16	3PL	84.71	7	20.77	174.87	Y
17	3PL	56.52	7	13.23	174.77	Y
18	3PL	68.82	7	16.52	174.78	Y
19	3PL	42.14	7	9.39	174.79	Y
20	3PL	129.41	7	32.72	174.80	Y
21	3PL	165.90	7	42.47	174.55	Y
22	3PL	241.20	7	62.59	174.80	Y
23	3PL	35.54	7	7.63	174.70	Y
24	3PL	166.17	7	42.54	174.23	Y
25	3PL	102.36	7	25.49	175.02	Y
26	3PL	82.91	7	20.29	175.07	Y
27	3PL	44.83	7	10.11	174.88	Y
28	3PL	105.50	7	26.33	174.84	Y
29	3PL	108.81	7	27.21	175.02	Y
30	3PL	53.79	7	12.50	174.88	Y
31	3PL	129.98	7	32.87	175.02	Y
32	3PL	225.33	7	58.35	174.67	Y
33	3PL	69.43	7	16.68	174.79	Y
34	3PL	59.47	7	14.02	174.92	Y
35	3PL	80.14	7	19.55	174.82	Y
36	3PL	117.23	7	29.46	175.01	Y
37	3PL	36.45	7	7.87	174.97	Y
38	3PL	181.72	7	46.70	174.91	Y
39	3PL	91.10	7	22.48	174.90	Y
40	3PL	90.66	7	22.36	174.85	Y
41	3PL	112.61	7	28.23	174.83	Y

Table O10. Mathematics Grade 6 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
42	3PL	83.49	7	20.44	174.85	Y
43	3PL	101.40	7	25.23	174.78	Y
44	3PL	58.46	7	13.75	174.90	Y
45	3PL	136.91	7	34.72	174.54	Y
46	3PL	125.69	7	31.72	174.83	Y
47	3PL	54.48	7	12.69	174.75	Y
48	3PL	150.16	7	38.26	173.84	Y
49	2PPC	266.98	16	44.37	173.92	Y
50	2PPC	386.43	16	65.48	174.66	Y
51	2PPC	572.20	16	98.32	174.32	Y
52	2PPC	428.66	16	72.95	174.09	Y
53	2PPC	372.80	16	63.07	173.29	Y
54	2PPC	413.29	16	70.23	174.27	Y
55	2PPC	308.41	25	40.08	173.29	Y
56	2PPC	712.16	25	97.18	174.30	Y
57	2PPC	745.19	25	101.85	172.86	Y
58	2PPC	519.36	25	69.91	173.47	Y

Table O11. Mathematics Grade 7 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	246.92	7	64.12	169.91	Y
2	3PL	179.66	7	46.15	169.87	Y
3	3PL	142.07	7	36.10	169.46	Y
4	3PL	43.75	7	9.82	169.77	Y
5	3PL	257.96	7	67.07	169.79	Y
6	3PL	39.14	7	8.59	169.75	Y
7	3PL	62.73	7	14.89	169.45	Y
8	3PL	39.75	7	8.75	169.66	Y
9	3PL	202.21	7	52.17	169.80	Y
10	3PL	36.51	7	7.89	169.76	Y
11	3PL	280.81	7	73.18	169.77	Y
12	3PL	107.76	7	26.93	169.71	Y
13	3PL	70.14	7	16.87	169.61	Y
14	3PL	49.00	7	11.23	169.76	Y
15	3PL	67.04	7	16.05	169.62	Y
16	3PL	68.18	7	16.35	169.72	Y
17	3PL	258.60	7	67.24	169.71	Y
18	3PL	104.73	7	26.12	169.69	Y
19	3PL	110.54	7	27.67	169.73	Y
20	3PL	298.32	7	77.86	169.41	Y

Table O11. Mathematics Grade 7 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
21	3PL	44.12	7	9.92	169.50	Y
22	3PL	100.53	7	25.00	169.43	Y
23	3PL	57.82	7	13.58	169.21	Y
24	3PL	86.79	7	21.32	168.97	Y
25	3PL	68.12	7	16.33	169.69	Y
26	3PL	191.70	7	49.36	169.89	Y
27	3PL	80.22	7	19.57	169.43	Y
28	3PL	164.91	7	42.20	169.90	Y
29	3PL	32.80	7	6.90	169.63	Y
30	3PL	180.62	7	46.40	169.88	Y
31	3PL	76.82	7	18.66	169.91	Y
32	3PL	458.44	7	120.65	169.81	Y
33	3PL	91.36	7	22.55	169.64	Y
34	3PL	658.94	7	174.24	169.85	N
35	3PL	69.88	7	16.81	169.85	Y
36	3PL	50.01	7	11.50	169.77	Y
37	3PL	110.57	7	27.68	169.80	Y
38	3PL	92.26	7	22.79	169.78	Y
39	3PL	74.95	7	18.16	169.71	Y
40	3PL	56.57	7	13.25	169.65	Y
41	3PL	114.10	7	28.62	169.62	Y
42	3PL	73.94	7	17.89	169.82	Y
43	3PL	36.51	7	7.89	169.80	Y
44	3PL	60.74	7	14.36	169.68	Y
45	3PL	94.87	7	23.48	169.72	Y
46	3PL	238.08	7	61.76	169.71	Y
47	3PL	223.77	7	57.93	169.64	Y
48	3PL	233.80	7	60.61	169.34	Y
49	2PPC	621.72	16	107.08	169.20	Y
50	2PPC	2592.71	16	455.50	168.72	N
51	2PPC	269.46	16	44.81	166.98	Y
52	2PPC	470.80	16	80.40	168.42	Y
53	2PPC	665.58	16	114.83	167.88	Y
54	2PPC	717.27	16	123.97	167.86	Y
55	2PPC	1084.36	25	149.82	168.93	Y
56	2PPC	385.84	25	51.03	168.26	Y
57	2PPC	457.88	25	61.22	167.23	Y
58	2PPC	1836.24	25	256.15	168.33	N

Table O12. Mathematics Grade 8 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	121.69	7	30.65	139.96	Y
2	3PL	54.24	7	12.63	140.02	Y
3	3PL	51.14	7	11.80	139.97	Y
4	3PL	88.70	7	21.83	139.91	Y
5	3PL	104.08	7	25.95	139.94	Y
6	3PL	47.60	7	10.85	140.00	Y
7	3PL	59.00	7	13.90	139.86	Y
8	3PL	55.32	7	12.91	139.99	Y
9	3PL	39.00	7	8.55	139.61	Y
10	3PL	141.65	7	35.99	139.98	Y
11	3PL	152.12	7	38.78	139.86	Y
12	3PL	52.09	7	12.05	139.91	Y
13	3PL	78.14	7	19.01	139.98	Y
14	3PL	136.31	7	34.56	139.98	Y
15	3PL	97.26	7	24.12	139.82	Y
16	3PL	225.29	7	58.34	139.95	Y
17	3PL	26.11	7	5.11	139.90	Y
18	3PL	204.76	7	52.85	139.97	Y
19	3PL	86.25	7	21.18	139.75	Y
20	3PL	122.28	7	30.81	139.95	Y
21	3PL	92.69	7	22.90	139.92	Y
22	3PL	108.70	7	27.18	139.91	Y
23	3PL	47.87	7	10.92	139.85	Y
24	3PL	71.11	7	17.13	139.66	Y
25	3PL	230.03	7	59.61	140.07	Y
26	3PL	118.46	7	29.79	139.80	Y
27	3PL	170.65	7	43.74	140.03	Y
28	3PL	99.12	7	24.62	139.79	Y
29	3PL	55.97	7	13.09	139.91	Y
30	3PL	51.98	7	12.02	139.95	Y
31	3PL	30.69	7	6.33	139.91	Y
32	3PL	49.22	7	11.28	139.91	Y
33	3PL	45.96	7	10.41	139.87	Y
34	3PL	146.09	7	37.17	139.96	Y
35	3PL	93.91	7	23.23	139.95	Y
36	3PL	103.08	7	25.68	139.72	Y
37	3PL	163.45	7	41.81	139.98	Y
38	3PL	22.82	7	4.23	139.91	Y
39	3PL	86.93	7	21.36	139.96	Y
40	3PL	20.51	7	3.61	139.82	Y
41	3PL	205.08	7	52.94	139.90	Y

Table O12. Mathematics Grade 8 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
42	3PL	183.84	7	47.26	139.95	Y
43	3PL	76.38	7	18.54	139.95	Y
44	3PL	81.83	7	20.00	139.99	Y
45	3PL	97.37	7	24.15	139.94	Y
46	3PL	59.90	7	14.14	139.90	Y
47	3PL	88.36	7	21.74	139.97	Y
48	3PL	38.63	7	8.45	139.88	Y
49	2PPC	553.23	16	94.97	139.11	Y
50	2PPC	456.70	16	77.91	137.11	Y
51	2PPC	261.25	16	43.35	135.88	Y
52	2PPC	435.49	16	74.16	135.37	Y
53	2PPC	332.01	16	55.86	136.88	Y
54	2PPC	522.63	16	89.56	136.03	Y
55	2PPC	384.86	25	50.89	137.59	Y
56	2PPC	637.23	25	86.58	137.51	Y
57	2PPC	1158.90	25	160.36	135.95	N
58	2PPC	597.55	25	80.97	137.01	Y

Table O13. ELA Grade 3 Operational Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.737	-1.321	0.030		
2	1	0.484	-2.040	0.007		
3	1	0.230	0.132	0.016		
4	1	0.658	-1.638	0.018		
5	1	0.604	-0.908	0.084		
6	1	0.468	0.836	0.162		
7	1	0.635	0.563	0.205		
8	1	0.468	-0.373	0.146		
9	1	0.596	0.072	0.140		
10	1	0.831	0.531	0.209		
11	1	0.422	1.973	0.188		
12	1	0.783	0.240	0.161		
13	1	0.686	1.035	0.226		
14	1	0.906	0.599	0.249		
15	1	0.570	0.407	0.128		
16	1	0.634	0.537	0.237		
17	1	1.080	-0.157	0.211		
18	1	0.646	0.505	0.181		
19	1	1.397	0.998	0.216		
20	1	0.859	0.190	0.222		

Table O13. ELA Grade 3 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
21	1	0.948	1.268	0.190		
22	1	0.918	0.482	0.207		
23	1	0.860	0.878	0.206		
24	1	0.714	-0.486	0.301		
25	1	0.629	-1.299	0.027		
26	1	0.699	1.419	0.282		
27	1	0.845	-0.074	0.192		
28	1	0.980	1.362	0.301		
29	1	0.923	0.783	0.221		
30	2	1.251	-3.021	0.532		
31	2	1.265	-1.678	1.154		
32	2	0.763	-1.028	1.150		
33	4	1.063	-1.949	-0.289	1.099	2.485
34	2	1.163	-2.067	1.802		
35	2	0.962	-1.255	1.566		
36	2	1.346	-1.483	1.573		
37	2	1.318	0.078	1.919		
38	2	1.331	-1.955	1.191		

Table O14. ELA Grade 4 Operational Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.485	-0.456	0.145		
2	1	0.925	-0.587	0.320		
3	1	0.640	-0.952	0.223		
4	1	0.620	1.052	0.288		
5	1	0.654	-0.655	0.148		
6	1	0.729	-0.342	0.139		
7	1	0.879	-0.512	0.230		
8	1	0.984	1.820	0.169		
9	1	0.571	-1.225	0.015		
10	1	0.715	0.383	0.180		
11	1	0.967	-0.541	0.186		
12	1	0.694	-0.981	0.063		
13	1	0.931	0.670	0.236		
14	1	0.286	-0.299	0.006		
15	1	0.714	-1.042	0.074		
16	1	0.310	-0.407	0.008		
17	1	0.849	0.466	0.179		
18	1	0.957	-0.585	0.157		
19	1	0.284	-0.326	0.049		

Table O14. ELA Grade 4 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
20	1	0.622	0.451	0.232		
21	1	0.615	0.639	0.223		
22	1	0.950	1.033	0.339		
23	1	0.278	1.061	0.067		
24	1	0.261	1.521	0.050		
25	1	0.102	-1.387	0.040		
26	1	0.417	-0.093	0.071		
27	1	0.471	0.433	0.135		
28	1	0.786	0.504	0.161		
29	1	0.466	3.463	0.212		
30	1	0.792	-0.083	0.181		
31	1	0.859	-0.665	0.209		
32	2	1.360	-2.603	0.671		
33	2	1.161	-3.036	0.676		
34	2	1.653	-1.936	1.353		
35	4	1.408	-2.059	-0.579	0.816	2.470
36	2	1.150	-1.703	1.743		
37	2	1.495	-2.344	1.764		
38	2	1.620	-2.538	1.508		
39	2	1.367	-1.199	1.581		
40	2	1.441	-2.660	1.067		
41	4	1.380	-2.762	-0.687	1.425	3.168

Table O15. ELA Grade 5 Operational Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.714	-2.153	0.151		
2	1	0.485	0.519	0.253		
3	1	0.517	-0.717	0.067		
4	1	0.804	-0.207	0.165		
5	1	0.663	0.716	0.340		
6	1	1.073	0.088	0.156		
7	1	0.537	-0.822	0.084		
8	1	0.697	0.850	0.196		
9	1	0.441	-1.458	0.005		
10	1	0.828	0.362	0.235		
11	1	0.628	1.182	0.206		
12	1	0.643	0.367	0.264		
13	1	0.893	-0.050	0.233		
14	1	0.823	-0.239	0.198		
15	1	0.486	0.221	0.217		

Table O15. ELA Grade 5 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par/alpha	b-par/step1	c-par / step2	step3	step4
16	1	0.673	0.876	0.198		
17	1	0.472	1.079	0.223		
18	1	1.093	-0.625	0.177		
19	1	0.818	0.572	0.163		
20	1	0.915	-1.471	0.146		
21	1	0.942	0.366	0.189		
22	1	0.835	-1.086	0.211		
23	1	0.772	0.993	0.296		
24	1	0.579	-0.082	0.140		
25	1	0.867	-0.839	0.146		
26	1	0.558	-1.234	0.014		
27	1	0.770	1.882	0.188		
28	1	1.119	0.511	0.188		
29	1	0.978	1.053	0.205		
30	1	1.002	-0.887	0.190		
31	1	0.868	0.371	0.211		
32	1	1.023	0.390	0.274		
33	1	0.435	-0.450	0.073		
34	1	0.747	-0.183	0.217		
35	1	0.898	0.242	0.203		
36	1	0.510	-0.158	0.471		
37	1	0.620	0.040	0.210		
38	1	1.192	2.390	0.237		
39	1	0.531	-0.389	0.195		
40	1	0.435	-1.022	0.124		
41	1	0.588	0.092	0.420		
42	1	0.410	-1.931	0.013		
43	2	1.137	-1.874	-0.591		
44	2	1.106	-2.566	-0.016		
45	2	1.482	-2.635	0.370		
46	4	1.366	-2.770	-0.963	0.864	2.569
47	2	1.080	-2.275	0.129		
48	2	0.875	-0.706	0.854		
49	2	1.059	-1.989	0.822		
50	2	0.892	0.221	1.164		
51	2	1.471	-2.477	0.690		
52	4	1.105	-1.760	-0.288	1.045	2.222

Table O16. ELA Grade 6 Operational Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.735	-0.329	0.065		
2	1	0.994	0.414	0.233		
3	1	0.686	-0.939	0.091		
4	1	0.658	-0.083	0.128		
5	1	0.959	-0.336	0.219		
6	1	0.793	0.012	0.173		
7	1	0.911	-1.321	0.173		
8	1	0.811	-0.335	0.150		
9	1	0.914	0.248	0.202		
10	1	0.917	-2.093	0.056		
11	1	0.530	-1.456	0.009		
12	1	0.380	0.963	0.115		
13	1	0.698	1.097	0.216		
14	1	0.450	-1.136	0.063		
15	1	0.822	1.045	0.193		
16	1	0.971	-0.694	0.213		
17	1	1.315	-0.949	0.220		
18	1	0.711	0.263	0.231		
19	1	0.898	-0.754	0.237		
20	1	0.532	0.429	0.188		
21	1	0.667	-0.565	0.009		
22	1	0.769	-0.427	0.115		
23	1	0.883	-0.413	0.182		
24	1	0.599	0.611	0.154		
25	1	1.288	0.368	0.251		
26	1	0.749	1.037	0.260		
27	1	0.722	0.679	0.227		
28	1	0.778	0.486	0.239		
29	1	0.966	0.756	0.164		
30	1	0.845	0.485	0.185		
31	1	1.152	-0.155	0.252		
32	1	0.940	1.328	0.220		
33	1	1.466	0.124	0.150		
34	1	0.416	-0.541	0.017		
35	1	0.805	-0.046	0.210		
36	1	0.783	-0.111	0.206		
37	1	0.603	-0.908	0.020		
38	1	0.384	-0.032	0.067		
39	1	0.637	-0.691	0.198		
40	1	1.090	0.393	0.224		
41	1	0.511	-1.006	0.010		

Table O16. ELA Grade 6 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
42	1	0.572	-0.671	0.169		
43	2	1.239	-2.835	-0.113		
44	2	1.271	-2.271	0.217		
45	2	1.063	-1.629	0.562		
46	4	1.137	-1.372	-0.171	0.906	1.983
47	2	1.103	-1.881	0.444		
48	2	1.244	-1.665	0.498		
49	2	1.297	-2.240	0.373		
50	2	1.364	-2.271	0.105		
51	2	1.209	-1.908	0.464		
52	4	1.139	-1.945	-0.525	0.904	1.868

Table O17. ELA Grade 7 Operational Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.830	-0.746	0.452		
2	1	0.968	0.880	0.328		
3	1	0.624	1.337	0.346		
4	1	0.686	0.287	0.248		
5	1	0.404	0.253	0.266		
6	1	0.853	-0.495	0.197		
7	1	0.380	0.588	0.041		
8	1	0.365	-2.035	0.012		
9	1	1.008	1.073	0.104		
10	1	1.566	-0.104	0.231		
11	1	0.411	0.717	0.102		
12	1	0.396	0.433	0.067		
13	1	0.603	1.142	0.129		
14	1	0.540	-0.678	0.197		
15	1	0.371	0.795	0.111		
16	1	0.997	-0.350	0.172		
17	1	0.441	0.499	0.085		
18	1	0.617	0.147	0.210		
19	1	0.881	1.200	0.192		
20	1	0.622	-0.660	0.073		
21	1	0.548	-0.426	0.079		
22	1	0.595	0.283	0.206		
23	1	0.929	1.192	0.116		
24	1	0.597	-0.124	0.141		
25	1	0.762	-1.116	0.073		
26	1	0.593	-0.834	0.014		

Table O17. ELA Grade 7 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
27	1	1.155	-0.251	0.159		
28	1	1.125	-0.606	0.300		
29	1	1.078	0.349	0.206		
30	1	0.970	0.576	0.184		
31	1	1.118	-0.617	0.252		
32	1	0.887	1.149	0.241		
33	1	0.557	0.749	0.149		
34	1	0.560	0.993	0.191		
35	1	0.641	-0.736	0.271		
36	1	0.684	-0.840	0.123		
37	1	0.711	0.043	0.227		
38	1	0.972	0.416	0.229		
39	1	0.822	0.161	0.333		
40	1	0.330	0.505	0.126		
41	1	0.894	0.558	0.183		
42	2	1.369	-2.864	0.134		
43	2	1.379	-2.598	-0.113		
44	2	1.124	-1.855	0.543		
45	4	1.313	-3.306	-1.540	0.476	2.080
46	2	1.253	-2.711	-0.051		
47	2	1.191	-2.529	-0.171		
48	2	1.275	-1.091	0.340		
49	2	1.374	-2.469	0.561		
50	2	1.457	-2.440	0.479		
51	4	1.253	-1.981	-0.585	0.626	1.637

Table O18. ELA Grade 8 Operational Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.457	-3.234	0.010		
2	1	1.113	0.331	0.222		
3	1	0.911	-0.746	0.270		
4	1	0.577	0.258	0.217		
5	1	1.027	0.695	0.166		
6	1	0.500	-0.077	0.196		
7	1	0.851	-0.530	0.227		
8	1	0.629	0.626	0.304		
9	1	0.666	-0.081	0.235		
10	1	0.760	1.441	0.275		
11	1	0.345	-2.459	0.068		
12	1	0.514	-1.082	0.092		

Table O18. ELA Grade 8 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
13	1	0.894	-0.031	0.156		
14	1	0.913	-1.324	0.204		
15	1	0.690	-0.105	0.218		
16	1	0.974	-1.323	0.229		
17	1	0.709	1.128	0.199		
18	1	0.960	0.095	0.289		
19	1	1.137	0.763	0.254		
20	1	0.451	1.026	0.111		
21	1	0.742	1.150	0.250		
22	1	1.005	0.110	0.203		
23	1	0.820	0.504	0.237		
24	1	1.039	-0.116	0.258		
25	1	0.427	0.022	0.102		
26	1	0.681	-0.005	0.118		
27	1	0.739	-0.080	0.128		
28	1	0.646	-0.486	0.145		
29	1	0.767	-0.056	0.122		
30	1	1.019	-0.028	0.129		
31	1	0.750	0.147	0.228		
32	1	0.589	0.825	0.229		
33	1	0.815	-0.032	0.230		
34	1	1.065	0.540	0.278		
35	1	0.770	0.358	0.196		
36	1	0.484	-0.194	0.143		
37	1	0.422	-2.196	0.034		
38	1	0.373	-1.322	0.175		
39	1	0.905	-0.505	0.278		
40	1	0.572	-0.244	0.164		
41	1	0.702	0.694	0.279		
42	1	0.891	-0.558	0.307		
43	2	1.130	-2.154	0.053		
44	2	1.243	-2.503	-0.033		
45	2	1.148	-2.147	0.284		
46	4	1.112	-1.883	-0.712	0.504	1.674
47	2	1.039	-2.198	-0.082		
48	2	0.862	-2.126	-0.783		
49	2	1.397	-2.620	-0.599		
50	2	1.337	-2.299	-0.066		
51	2	1.085	-2.114	0.025		
52	4	1.140	-2.574	-1.597	-0.035	1.534

Table O19. Mathematics Grade 3 Operational Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.616	-2.393	0.019	
2	1	0.639	-2.182	0.010	
3	1	0.909	-2.016	0.234	
4	1	0.383	-1.685	0.025	
5	1	0.614	-2.254	0.020	
6	1	1.198	0.207	0.111	
7	1	1.425	0.236	0.068	
8	1	0.952	-0.977	0.052	
9	1	0.986	-1.151	0.129	
10	1	0.820	-1.490	0.005	
11	1	0.780	-0.783	0.003	
12	1	0.415	1.788	0.209	
13	1	0.755	-0.131	0.198	
14	1	1.668	0.280	0.109	
15	1	1.254	0.077	0.086	
16	1	0.676	-0.315	0.136	
17	1	1.000	-0.082	0.175	
18	1	0.847	0.494	0.334	
19	1	1.006	0.482	0.120	
20	1	0.732	-0.755	0.153	
21	1	1.113	-1.367	0.251	
22	1	0.875	-2.604	0.041	
23	1	0.886	-0.865	0.199	
24	1	0.990	-0.789	0.184	
25	1	0.904	-0.786	0.058	
26	1	1.273	0.681	0.477	
27	1	0.628	0.712	0.162	
28	1	1.079	0.457	0.119	
29	1	1.321	-0.029	0.145	
30	1	0.820	-1.593	0.027	
31	1	1.285	0.733	0.198	
32	1	0.972	-0.234	0.236	
33	1	0.870	-0.275	0.176	
34	1	0.694	-1.671	0.007	
35	1	1.258	0.042	0.243	
36	1	0.671	-1.218	0.220	
37	1	0.897	-0.189	0.190	
38	1	0.996	0.233	0.075	
39	1	0.665	-0.043	0.266	
40	1	0.510	-2.074	0.026	
41	1	1.433	0.263	0.216	

Table O19. Mathematics Grade 3 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
42	2	0.917	0.204	0.265	
43	2	1.136	0.376	0.266	
44	2	0.988	-0.967	0.217	
45	2	0.787	2.541	-1.122	
46	2	0.813	-0.473	0.443	
47	3	0.933	-1.316	0.850	0.494
48	3	1.069	0.463	0.745	1.029
49	3	1.028	0.485	-0.511	0.626

Table O20. Mathematics Grade 4 Operational Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	1.450	-0.935	0.216	
2	1	0.949	-0.545	0.291	
3	1	0.680	-0.048	0.230	
4	1	0.912	-0.644	0.173	
5	1	0.941	-1.373	0.543	
6	1	0.817	-1.025	0.360	
7	1	0.791	0.211	0.237	
8	1	1.040	0.877	0.138	
9	1	0.715	-0.789	0.276	
10	1	0.887	0.469	0.295	
11	1	1.013	0.298	0.068	
12	1	0.953	-0.615	0.234	
13	1	0.914	-0.196	0.100	
14	1	1.055	0.082	0.175	
15	1	0.952	-0.172	0.142	
16	1	0.882	-0.795	0.247	
17	1	0.792	-0.009	0.121	
18	1	0.850	0.464	0.173	
19	1	1.111	-0.125	0.158	
20	1	1.109	0.232	0.151	
21	1	0.631	-1.630	0.015	
22	1	0.735	-1.590	0.083	
23	1	1.271	-0.202	0.029	
24	1	1.076	-0.159	0.287	
25	1	0.851	0.520	0.163	
26	1	0.970	-1.458	0.020	
27	1	1.323	-0.155	0.228	
28	1	0.912	1.168	0.243	
29	1	1.193	0.748	0.130	

Table O20. Mathematics Grade 4 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
30	1	0.563	-1.198	0.007	
31	1	1.196	0.043	0.315	
32	1	0.997	0.533	0.075	
33	1	1.220	-0.295	0.070	
34	1	0.654	-0.515	0.039	
35	1	1.082	0.483	0.182	
36	1	1.276	0.250	0.228	
37	1	0.856	0.675	0.174	
38	1	0.872	0.504	0.053	
39	1	0.685	-0.090	0.115	
40	1	1.222	0.417	0.105	
41	1	0.938	-0.904	0.267	
42	1	0.762	-1.028	0.176	
43	2	0.772	0.132	-0.569	
44	2	1.437	0.245	0.132	
45	2	0.967	0.649	-1.789	
46	2	0.770	-0.107	-1.006	
47	2	1.346	-0.168	-0.102	
48	2	1.523	-0.221	-0.445	
49	3	1.308	-2.287	1.389	0.219
50	3	1.105	0.493	-0.092	1.207
51	3	0.817	0.970	-0.267	-0.120
52	3	0.971	-0.652	-0.615	1.052

Table O21. Mathematics Grade 5 Operational Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.637	-1.401	0.106	
2	1	0.811	-0.329	0.359	
3	1	0.900	-1.321	0.017	
4	1	1.190	0.058	0.500	
5	1	1.150	0.526	0.288	
6	1	1.159	-1.102	0.141	
7	1	1.322	0.953	0.165	
8	1	0.983	-1.013	0.103	
9	1	1.098	0.113	0.267	
10	1	1.210	0.688	0.202	
11	1	0.643	0.695	0.190	
12	1	1.510	0.919	0.140	
13	1	0.963	-0.159	0.180	
14	1	1.035	-0.119	0.292	

Table O21. Mathematics Grade 5 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
15	1	1.626	0.701	0.280	
16	1	0.905	-1.118	0.033	
17	1	1.148	-0.262	0.319	
18	1	1.078	0.981	0.339	
19	1	1.241	0.604	0.284	
20	1	1.518	0.404	0.091	
21	1	0.841	0.150	0.216	
22	1	1.294	0.304	0.465	
23	1	0.894	-1.570	0.266	
24	1	1.056	-0.732	0.219	
25	1	1.194	-0.417	0.127	
26	1	1.078	0.124	0.296	
27	1	0.917	0.210	0.234	
28	1	0.297	-1.666	0.076	
29	1	1.561	0.032	0.154	
30	1	0.558	0.428	0.272	
31	1	0.929	0.444	0.337	
32	1	1.126	1.489	0.196	
33	1	0.590	-1.826	0.027	
34	1	1.375	1.229	0.169	
35	1	1.255	-0.094	0.327	
36	1	0.946	0.166	0.211	
37	1	1.371	0.071	0.214	
38	1	1.300	0.797	0.254	
39	1	0.931	-0.057	0.257	
40	1	0.992	1.557	0.226	
41	1	0.886	-0.673	0.130	
42	1	1.217	-0.398	0.085	
43	2	0.781	-1.440	-0.974	
44	2	1.229	-2.064	0.656	
45	2	0.955	0.726	0.684	
46	2	1.068	-0.047	-2.171	
47	2	1.362	0.091	0.414	
48	2	1.705	-0.057	1.103	
49	3	0.992	0.862	1.632	0.779
50	3	0.927	1.149	1.545	0.651
51	3	1.206	1.346	2.500	1.123
52	3	1.105	0.200	1.520	0.407

Table O22. Mathematics Grade 6 Operational Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	1.242	-1.008	0.034	
2	1	0.926	-0.702	0.104	
3	1	0.989	-0.049	0.087	
4	1	0.967	-0.944	0.125	
5	1	1.160	-1.118	0.148	
6	1	1.118	0.306	0.118	
7	1	0.767	0.903	0.374	
8	1	0.813	-1.372	0.149	
9	1	0.571	0.663	0.225	
10	1	0.973	0.379	0.224	
11	1	1.011	-0.173	0.091	
12	1	0.497	0.911	0.255	
13	1	1.000	0.484	0.260	
14	1	0.963	-0.635	0.080	
15	1	1.005	1.247	0.200	
16	1	0.911	0.523	0.291	
17	1	0.809	-0.083	0.274	
18	1	0.795	1.283	0.208	
19	1	1.303	-0.323	0.329	
20	1	1.209	-0.230	0.365	
21	1	0.891	0.397	0.292	
22	1	0.730	-2.256	0.018	
23	1	0.647	0.590	0.259	
24	1	1.410	-0.837	0.142	
25	1	0.877	-1.524	0.171	
26	1	0.733	-1.538	0.250	
27	1	0.819	-0.532	0.288	
28	1	1.514	0.123	0.101	
29	1	1.068	-0.485	0.175	
30	1	0.764	-0.334	0.115	
31	1	1.301	0.830	0.100	
32	1	1.726	1.234	0.236	
33	1	0.944	0.558	0.285	
34	1	0.905	0.261	0.227	
35	1	1.267	0.731	0.328	
36	1	0.841	-0.624	0.098	
37	1	1.003	-0.776	0.269	
38	1	1.263	0.440	0.149	
39	1	0.879	0.356	0.169	
40	1	1.111	-0.027	0.166	
41	1	0.639	0.097	0.403	

Table O22. Mathematics Grade 6 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
42	1	1.290	0.845	0.231	
43	1	1.409	0.872	0.191	
44	1	0.774	-0.170	0.231	
45	1	1.439	1.186	0.287	
46	1	1.798	0.435	0.195	
47	1	0.894	-0.725	0.274	
48	1	0.479	-1.032	0.041	
49	2	1.086	0.047	-0.640	
50	2	1.201	-0.870	1.428	
51	2	1.028	0.733	1.653	
52	2	1.050	0.811	-0.093	
53	2	0.993	0.502	-1.439	
54	2	1.069	-0.220	-0.761	
55	3	0.849	1.250	1.633	1.002
56	3	1.340	-0.011	0.636	1.876
57	3	1.267	1.971	2.065	0.860
58	3	0.924	-0.578	-0.113	-1.292

Table O23. Mathematics Grade 7 Operational Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.510	-1.879	0.124	
2	1	0.770	-0.627	0.130	
3	1	0.919	-0.513	0.209	
4	1	0.940	0.364	0.188	
5	1	0.675	1.505	0.288	
6	1	1.287	1.505	0.287	
7	1	1.075	0.983	0.149	
8	1	1.300	1.023	0.238	
9	1	0.754	-0.668	0.072	
10	1	0.825	0.666	0.302	
11	1	1.787	1.260	0.145	
12	1	1.200	-0.138	0.231	
13	1	1.248	0.722	0.091	
14	1	1.233	1.073	0.201	
15	1	1.314	0.443	0.250	
16	1	0.893	0.275	0.451	
17	1	0.992	-0.296	0.042	
18	1	0.912	0.116	0.356	
19	1	0.910	-0.488	0.258	
20	1	1.490	1.369	0.205	

Table O23. Mathematics Grade 7 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
21	1	0.659	0.399	0.100	
22	1	0.457	1.599	0.258	
23	1	1.045	0.702	0.330	
24	1	1.151	0.484	0.194	
25	1	1.193	0.147	0.360	
26	1	1.368	-0.745	0.239	
27	1	1.122	0.550	0.277	
28	1	0.640	-0.770	0.180	
29	1	1.209	0.230	0.217	
30	1	0.629	-1.589	0.018	
31	1	1.480	0.688	0.115	
32	1	2.147	1.387	0.264	
33	1	1.633	1.354	0.281	
34	1	2.624	1.286	0.096	
35	1	1.325	0.582	0.167	
36	1	1.187	0.175	0.202	
37	1	1.239	0.545	0.231	
38	1	1.277	-0.528	0.331	
39	1	1.276	0.410	0.163	
40	1	1.166	0.988	0.382	
41	1	1.698	0.478	0.242	
42	1	0.799	0.024	0.144	
43	1	0.858	0.281	0.248	
44	1	1.182	0.822	0.280	
45	1	1.447	-0.313	0.321	
46	1	1.052	-0.572	0.124	
47	1	1.891	0.432	0.187	
48	1	0.890	-1.016	0.164	
49	2	1.147	-0.742	-0.531	
50	2	1.319	-0.117	-0.159	
51	2	1.561	1.452	1.214	
52	2	1.595	-0.116	0.500	
53	2	1.524	0.248	0.074	
54	2	1.163	-0.012	0.288	
55	3	1.441	0.999	1.344	1.534
56	3	0.942	0.432	-1.165	2.279
57	3	1.399	0.629	1.990	1.063
58	3	0.935	-0.025	0.002	-0.258

Table O24. Mathematics Grade 8 Operational Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.726	-0.990	0.176	
2	1	1.266	-0.167	0.415	
3	1	0.549	0.205	0.158	
4	1	0.865	0.365	0.216	
5	1	0.867	-0.555	0.248	
6	1	0.862	0.092	0.242	
7	1	1.322	0.909	0.165	
8	1	0.741	-0.033	0.218	
9	1	1.127	0.537	0.215	
10	1	0.494	-2.050	0.019	
11	1	1.274	0.853	0.144	
12	1	1.282	0.584	0.232	
13	1	0.737	-0.330	0.198	
14	1	1.129	-0.530	0.172	
15	1	1.588	0.648	0.135	
16	1	1.002	-1.112	0.201	
17	1	0.829	0.365	0.234	
18	1	1.902	1.213	0.350	
19	1	1.152	0.643	0.265	
20	1	0.698	-0.714	0.168	
21	1	0.384	-0.353	0.196	
22	1	0.826	-0.403	0.142	
23	1	0.434	0.567	0.086	
24	1	0.754	-0.137	0.268	
25	1	0.697	-2.298	0.025	
26	1	0.825	0.007	0.149	
27	1	0.529	-0.845	0.011	
28	1	1.719	0.722	0.160	
29	1	1.230	1.822	0.383	
30	1	1.283	0.091	0.257	
31	1	0.675	0.849	0.292	
32	1	0.929	0.514	0.202	
33	1	1.271	0.061	0.356	
34	1	0.575	-0.687	0.211	
35	1	0.698	-0.657	0.095	
36	1	1.984	0.833	0.229	
37	1	0.507	-1.241	0.017	
38	1	0.666	0.976	0.247	
39	1	0.752	-0.585	0.205	
40	1	0.862	1.072	0.190	
41	1	0.789	-0.017	0.187	

Table O24. Mathematics Grade 8 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
42	1	0.906	-1.010	0.152	
43	1	0.754	0.053	0.157	
44	1	0.786	-0.215	0.295	
45	1	1.420	0.384	0.286	
46	1	1.253	0.367	0.271	
47	1	1.658	0.194	0.242	
48	1	0.496	-0.487	0.208	
49	2	1.570	-0.968	-0.582	
50	2	1.649	-0.084	1.817	
51	2	1.625	0.671	1.366	
52	2	0.388	0.811	0.493	
53	2	1.596	1.370	1.756	
54	2	1.190	0.432	0.808	
55	3	1.145	-0.610	0.397	1.486
56	3	1.329	-0.118	-0.270	0.593
57	3	0.659	-0.802	1.083	1.684
58	3	1.271	-0.415	-0.219	-0.064

Appendix P: Derivation and Estimation of Classification Consistency and Accuracy

Classification Consistency

Assume that θ is a single latent trait measured by a test and denote Φ as a latent random variable. When a test X consists of K items and its maximum number correct score is N , the marginal probability of the number correct (NC) score x is

$$P(X = x) = \int P(X = x | \Phi = \theta) g(\theta) d\theta, \quad x = 0, 1, \dots, N$$

where

$g(\theta)$ is the density of θ .

In this report, the marginal distribution $P(X = x)$ is denoted as $f(x)$, and the conditional error distribution $P(X = x | \Phi = \theta)$ is denoted as $f(x | \theta)$. It is assumed that examinees are classified into one of H mutually exclusive categories on the basis of predetermined $H - 1$ observed score cutoffs, C_1, C_2, \dots, C_{H-1} . Let L_h represent the h th category into which examinees with $C_{h-1} \leq X < C_h$ are classified. $C_0 = 0$ and $C_H =$ the maximum number-correct score plus one. Then, the conditional and marginal probabilities of each category classification are as follows:

$$P(X \in L_h | \theta) = \sum_{x=C_{h-1}}^{C_h-1} f(x | \theta), \quad h = 1, 2, \dots, H$$

$$P(X \in L_h) = \int \sum_{x=C_{h-1}}^{C_h-1} f(x | \theta) g(\theta) d\theta, \quad h = 1, 2, \dots, H$$

Because obtaining test scores from two independent administrations of New York State tests was not feasible due to item release after each OP administration, a psychometric model was used to obtain the estimated classification consistency indices using test scores from a single administration. Based on the psychometric model, a symmetric H -by- H contingency table can be constructed. The elements of the H -by- H contingency table consist of the joint probabilities of the row and column observed category classifications.

That two administrations are independent implies that if X_1 and X_2 represent the raw score random variables on the two administrations, then, conditioned on θ , X_1 and X_2 are independent and identically distributed. Consequently, the conditional bivariate distribution of X_1 and X_2 is

$$f(x_1, x_2 | \theta) = f(x_1 | \theta) f(x_2 | \theta)$$

The marginal bivariate distribution of X_1 and X_2 can be expressed as follows:

$$f(x_1, x_2) = \int f(x_1, x_2 | \theta) f(\theta) d\theta$$

Consistent classification means that both X_1 and X_2 fall in the same category. The conditional probability of falling in the same category on the two administrations is

$$P(X_1 \in L_h, X_2 \in L_h | \theta) = \left[\sum_{x_1=C_{h-1}}^{C_{h+1}} f(x_1 | \theta) \right]^2, \quad h = 1, 2, \dots, H$$

The agreement index P , conditional on theta, is obtained by

$$P(\theta) = \sum_{h=1}^H P(X_1 \in L_h, X_2 \in L_h | \theta)$$

The agreement index (classification consistency) can be computed as

$$P = \int P(\theta) g(\theta) d(\theta)$$

The probability of consistent classification by chance, P_c , is the sum of squared marginal probabilities of each category classification.

$$P_c = \sum_{h=1}^H P(X_1 \in L_h) P(X_2 \in L_h) = \sum_{h=1}^H [P(X_1 \in L_h)]^2$$

Then, Kappa (Cohen, 1960) is

$$k = \frac{P - P_c}{1 - P_c}$$

Classification Accuracy

Let Γ_w denote true category. When an examinee has an observed score, $x \in L_h$ ($h = 1, 2, \dots, H$), and a latent score, $\theta \in \Gamma_w$ ($w = 1, 2, \dots, H$), an accurate classification is made when $h = w$. The conditional probability of accurate classification is

$$\gamma(\theta) = P(X \in L_w | \theta),$$

where

w is the category such that $\theta \in \Gamma_w$

Lee (2008) thoroughly discusses this IRT method for estimating decision indices, including the computational method used to estimate the results when integrating across the latent variable, θ .

Estimating Classification Indices

The classification consistency and accuracy estimates were obtained using an open-source software program, IRT-CLASS v2.0 (Lee & Kolen, 2006). Below is a brief description of the files that are used and their purpose. (See the IRT-CLASS v2.0 manual for complete instructions.)

Files needed:

- Raw-to-Scale score conversion file
 - a. Contains the raw-to-scale score conversions
 - b. This is used to provide both raw and scale score classification estimates, which is useful when the raw-to-scale score transformation is not one-to-one.
- Cut score file
 - a. Contains the cut scores to be used
 - b. Results are provided for all cut scores simultaneously (all performance levels), as well as the estimates based on each of the cut scores separately (Level 3 only).
- Item parameter file
 - a. This contains the IRT model used and item parameter estimates.
 - b. This information is used when calculating the classification indices.
- Theta file
 - a. Contains the theta distribution in terms of quadrature points
 - b. The theta and the item parameter files are used to solve the integrals mentioned above.
- Control card
 - a. This is used to run the program.
 - b. Identifies the names of the four files above and gives a name to the output file

Appendix Q: Raw Score-to-Scale Score and Scale Score Frequency Tables

Tables R1–R12 show the raw score-to-scale score conversion tables, while Tables R13–R24 show the scale score distributions, by frequency (n-count), percent, cumulative frequency, and cumulative percent. The data in the tables include all students with valid scores.

Table Q1. ELA Grade 3 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	148	51	25	300	10
1	156	45	26	304	10
2	164	40	27	307	10
3	172	35	28	311	10
4	180	31	29	314	10
5	188	27	30	320	10
6	196	24	31	321	10
7	204	22	32	324	10
8	212	20	33	327	10
9	220	18	34	331	10
10	228	17	35	334	10
11	235	16	36	338	10
12	242	15	37	341	10
13	248	14	38	345	10
14	254	14	39	349	10
15	259	13	40	354	11
16	264	13	41	358	11
17	269	12	42	363	12
18	273	12	43	369	13
19	278	12	44	376	14
20	282	12	45	384	16
21	286	11	46	395	19
22	291	11	47	403	21
23	293	11	48	411	24
24	297	11	49	419	27

Table Q2. ELA Grade 4 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	138	60	28	292	9
1	146	52	29	296	9
2	154	45	30	299	9
3	162	39	31	302	9
4	170	34	32	305	9
5	178	30	33	308	9
6	186	26	34	311	9
7	194	23	35	315	9
8	202	20	36	320	9
9	210	18	37	321	9
10	218	16	38	325	9
11	226	14	39	328	10
12	232	13	40	332	10
13	237	13	41	335	10
14	242	12	42	339	10
15	247	11	43	343	10
16	251	11	44	347	11
17	255	11	45	351	11
18	259	10	46	356	11
19	263	10	47	361	12
20	266	10	48	367	13
21	270	10	49	374	14
22	273	10	50	381	15
23	276	10	51	391	18
24	280	10	52	399	20
25	283	9	53	407	23
26	287	9	54	415	26
27	289	9	55	423	29

Table Q3. ELA Grade 5 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	105	90	34	289	9
1	113	79	35	292	9
2	121	69	36	294	9
3	129	61	37	297	9
4	137	53	38	299	9
5	145	47	39	302	9
6	153	41	40	304	9
7	161	36	41	307	8
8	169	32	42	309	8
9	177	28	43	312	9
10	185	24	44	314	9
11	193	22	45	317	9
12	201	19	46	320	9
13	209	17	47	322	9
14	217	16	48	325	9
15	225	14	49	327	9
16	230	13	50	330	9
17	235	13	51	333	9
18	239	12	52	337	9
19	243	12	53	340	10
20	247	11	54	346	10
21	251	11	55	347	10
22	254	11	56	351	11
23	258	10	57	355	11
24	261	10	58	359	11
25	264	10	59	364	12
26	267	10	60	370	13
27	270	10	61	376	14
28	273	10	62	384	15
29	276	9	63	393	17
30	278	9	64	401	19
31	281	9	65	409	22
32	284	9	66	417	25
33	286	9			

Table Q4. ELA Grade 6 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	118	80	34	288	8
1	126	69	35	290	8
2	134	60	36	293	8
3	142	52	37	295	8
4	150	45	38	297	8
5	158	39	39	299	8
6	166	34	40	301	8
7	174	29	41	304	8
8	183	25	42	306	8
9	191	22	43	308	8
10	199	19	44	310	8
11	207	17	45	313	8
12	215	15	46	315	8
13	223	14	47	317	8
14	228	13	48	320	8
15	233	12	49	322	8
16	237	12	50	325	8
17	241	11	51	327	8
18	245	11	52	330	9
19	248	11	53	333	9
20	252	10	54	338	9
21	255	10	55	340	9
22	258	10	56	343	10
23	261	9	57	347	10
24	264	9	58	351	11
25	266	9	59	356	11
26	269	9	60	361	12
27	272	9	61	367	13
28	274	9	62	374	15
29	277	9	63	383	17
30	279	8	64	396	21
31	283	8	65	404	24
32	284	8	66	412	28
33	286	8			

Table Q5. ELA Grade 7 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	113	87	33	288	8
1	121	75	34	291	8
2	129	65	35	293	8
3	137	56	36	295	8
4	145	49	37	298	8
5	153	42	38	300	8
6	161	36	39	303	8
7	169	31	40	305	8
8	177	27	41	307	8
9	185	24	42	310	8
10	193	21	43	312	8
11	201	18	44	314	8
12	209	16	45	318	8
13	217	15	46	320	8
14	225	14	47	322	8
15	231	13	48	325	8
16	235	12	49	328	9
17	240	12	50	330	9
18	244	11	51	333	9
19	248	11	52	336	9
20	252	11	53	340	9
21	255	10	54	343	10
22	258	10	55	347	10
23	262	10	56	351	11
24	265	10	57	355	11
25	268	9	58	360	12
26	270	9	59	366	13
27	273	9	60	372	14
28	276	9	61	380	16
29	278	9	62	390	18
30	281	9	63	398	21
31	283	9	64	406	24
32	287	8	65	414	27

Table Q6. ELA Grade 8 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	97	80	34	284	9
1	105	71	35	285	9
2	113	63	36	287	9
3	120	57	37	289	8
4	128	51	38	292	8
5	136	45	39	294	8
6	144	40	40	296	8
7	152	36	41	299	8
8	160	32	42	301	8
9	168	29	43	303	8
10	176	25	44	306	8
11	184	23	45	308	8
12	192	20	46	311	8
13	200	18	47	313	8
14	207	17	48	316	8
15	215	15	49	318	8
16	223	14	50	321	9
17	228	13	51	324	9
18	233	13	52	326	9
19	237	12	53	329	9
20	241	12	54	333	9
21	245	11	55	336	10
22	248	11	56	339	10
23	252	11	57	343	10
24	255	10	58	347	11
25	258	10	59	352	11
26	261	10	60	357	12
27	264	10	61	363	13
28	267	10	62	371	15
29	269	9	63	380	17
30	272	9	64	393	21
31	275	9	65	401	24
32	277	9	66	409	28
33	280	9			

Table Q7. Mathematics Grade 3 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	128	67	31	290	8
1	136	58	32	292	8
2	144	50	33	294	7
3	152	43	34	296	7
4	160	38	35	298	7
5	168	33	36	301	7
6	176	29	37	303	7
7	184	25	38	305	7
8	192	22	39	307	7
9	200	20	40	309	7
10	208	18	41	311	7
11	216	16	42	314	7
12	224	15	43	315	7
13	230	14	44	318	7
14	236	13	45	320	7
15	240	12	46	322	8
16	245	12	47	325	8
17	249	11	48	327	8
18	253	11	49	330	8
19	256	11	50	333	9
20	260	10	51	337	9
21	263	10	52	340	10
22	266	10	53	344	10
23	269	9	54	349	11
24	272	9	55	354	12
25	275	9	56	361	14
26	278	9	57	370	17
27	280	8	58	382	21
28	283	8	59	390	25
29	285	8	60	399	29
30	288	8			

Table Q8. Mathematics Grade 4 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	134	97	34	292	7
1	143	82	35	294	7
2	151	71	36	296	7
3	159	61	37	297	7
4	167	53	38	299	6
5	175	45	39	301	6
6	184	38	40	302	6
7	192	33	41	304	6
8	200	28	42	306	6
9	208	24	43	307	6
10	216	21	44	309	7
11	224	18	45	311	7
12	232	16	46	313	7
13	238	14	47	314	7
14	243	13	48	316	7
15	247	12	49	318	7
16	251	11	50	320	7
17	255	11	51	322	7
18	258	10	52	324	7
19	261	10	53	327	7
20	264	9	54	329	8
21	266	9	55	332	8
22	269	8	56	334	8
23	271	8	57	337	8
24	273	8	58	341	9
25	275	8	59	344	10
26	278	8	60	348	10
27	280	7	61	353	11
28	281	7	62	359	13
29	283	7	63	366	14
30	285	7	64	377	18
31	287	7	65	396	27
32	289	7	66	404	31
33	291	7			

Table Q9. Mathematics Grade 5 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	125	103	34	302	7
1	133	92	35	304	7
2	141	81	36	306	7
3	149	72	37	308	7
4	157	63	38	310	7
5	165	55	39	311	7
6	173	48	40	313	7
7	182	42	41	315	7
8	190	36	42	317	7
9	198	31	43	319	7
10	206	27	44	321	7
11	214	24	45	323	7
12	222	20	46	325	7
13	230	18	47	327	7
14	238	15	48	329	7
15	244	14	49	331	7
16	249	13	50	333	7
17	254	12	51	335	7
18	258	11	52	338	7
19	262	11	53	340	7
20	266	10	54	342	7
21	269	10	55	346	8
22	272	10	56	347	8
23	275	9	57	350	8
24	278	9	58	353	8
25	281	9	59	357	9
26	283	8	60	360	10
27	286	8	61	365	11
28	288	8	62	370	12
29	291	8	63	377	14
30	294	7	64	386	17
31	295	7	65	394	21
32	297	7	66	402	25
33	299	7			

Table Q10. Mathematics Grade 6 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	121	148	37	297	7
1	129	126	38	299	7
2	137	108	39	301	7
3	145	92	40	303	7
4	153	79	41	305	7
5	161	68	42	306	7
6	169	58	43	308	7
7	178	49	44	310	7
8	186	42	45	312	7
9	194	36	46	314	7
10	202	31	47	316	7
11	210	26	48	318	7
12	218	22	49	320	7
13	226	19	50	322	7
14	234	16	51	324	7
15	240	14	52	326	7
16	245	13	53	328	7
17	249	12	54	330	7
18	253	11	55	332	7
19	256	10	56	334	7
20	260	10	57	336	7
21	263	9	58	340	7
22	265	9	59	341	7
23	268	9	60	343	7
24	270	8	61	346	8
25	273	8	62	349	8
26	275	8	63	352	8
27	277	8	64	355	9
28	279	8	65	358	9
29	282	8	66	363	10
30	284	7	67	367	11
31	286	7	68	373	13
32	288	7	69	381	15
33	290	7	70	392	19
34	292	7	71	400	22
35	293	7	72	408	26
36	295	7			

Table Q11. Mathematics Grade 7 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	134	154	37	306	6
1	142	130	38	308	6
2	149	113	39	309	6
3	157	96	40	311	6
4	165	82	41	312	6
5	173	70	42	314	6
6	181	60	43	315	6
7	189	52	44	317	6
8	196	45	45	318	6
9	204	39	46	320	6
10	212	33	47	322	6
11	220	28	48	323	6
12	228	24	49	324	6
13	235	21	50	326	6
14	244	17	51	327	6
15	251	15	52	329	6
16	257	13	53	330	6
17	261	12	54	332	6
18	265	11	55	334	6
19	269	10	56	335	6
20	272	9	57	337	6
21	275	9	58	339	6
22	278	9	59	341	6
23	280	8	60	343	6
24	282	8	61	345	6
25	285	8	62	348	6
26	287	7	63	349	6
27	289	7	64	352	7
28	291	7	65	355	7
29	293	7	66	358	8
30	295	7	67	362	9
31	296	7	68	367	10
32	298	6	69	374	12
33	300	6	70	384	16
34	301	6	71	392	20
35	303	6	72	400	24
36	305	6			

Table Q12. Mathematics Grade 8 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	117	114	37	301	7
1	125	100	38	302	7
2	133	88	39	304	7
3	141	77	40	306	6
4	149	68	41	307	6
5	157	60	42	309	6
6	164	53	43	311	6
7	172	47	44	313	6
8	180	42	45	314	6
9	188	37	46	316	6
10	196	33	47	318	6
11	204	29	48	319	6
12	212	26	49	322	6
13	220	23	50	323	6
14	228	20	51	325	6
15	236	17	52	326	6
16	243	15	53	328	6
17	248	14	54	330	6
18	253	13	55	332	6
19	257	12	56	334	7
20	261	11	57	336	7
21	265	10	58	339	7
22	268	10	59	341	7
23	271	10	60	344	7
24	274	9	61	346	8
25	276	9	62	349	8
26	279	8	63	353	9
27	281	8	64	357	9
28	283	8	65	361	10
29	287	8	66	366	11
30	288	8	67	372	13
31	290	7	68	380	15
32	292	7	69	391	19
33	293	7	70	399	22
34	295	7	71	407	26
35	297	7	72	415	29
36	299	7			

Table Q13. ELA Grade 3 Scale Score Frequency Distribution

Scale Score	N-Count	%	Cumulative	
			N-Count	%
148	20	0.01%	20	0.01%
156	28	0.01%	48	0.02%
164	65	0.03%	113	0.06%
172	167	0.08%	280	0.14%
180	352	0.18%	632	0.32%
188	657	0.33%	1,289	0.64%
196	1,033	0.52%	2,322	1.16%
204	1,467	0.73%	3,789	1.89%
212	1,885	0.94%	5,674	2.83%
220	2,401	1.20%	8,075	4.03%
228	2,767	1.38%	10,842	5.41%
235	3,163	1.58%	14,005	6.98%
242	3,709	1.85%	17,714	8.83%
248	4,237	2.11%	21,951	10.95%
254	4,728	2.36%	26,679	13.31%
259	5,327	2.66%	32,006	15.96%
264	5,940	2.96%	37,946	18.93%
269	6,603	3.29%	44,549	22.22%
273	6,976	3.48%	51,525	25.70%
278	7,262	3.62%	58,787	29.32%
282	7,641	3.81%	66,428	33.13%
286	7,898	3.94%	74,326	37.07%
291	7,974	3.98%	82,300	41.05%
293	8,045	4.01%	90,345	45.06%
297	8,184	4.08%	98,529	49.14%
300	8,046	4.01%	106,575	53.15%
304	7,935	3.96%	114,510	57.11%
307	7,827	3.90%	122,337	61.01%
311	7,631	3.81%	129,968	64.82%
314	7,190	3.59%	137,158	68.41%
320	7,039	3.51%	144,197	71.92%
321	6,721	3.35%	150,918	75.27%
324	6,513	3.25%	157,431	78.52%
327	6,075	3.03%	163,506	81.55%
331	5,594	2.79%	169,100	84.34%
334	5,128	2.56%	174,228	86.90%
338	4,785	2.39%	179,013	89.28%
341	4,289	2.14%	183,302	91.42%
345	3,754	1.87%	187,056	93.29%
349	3,390	1.69%	190,446	94.98%

Table Q13. ELA Grade 3 Scale Score Frequency Distribution (cont.)

Scale Score	N-Count	%	Cumulative	
			N-Count	%
354	2,772	1.38%	193,218	96.37%
358	2,260	1.13%	195,478	97.49%
363	1,768	0.88%	197,246	98.38%
369	1,272	0.63%	198,518	99.01%
376	866	0.43%	199,384	99.44%
384	568	0.28%	199,952	99.72%
395	335	0.17%	200,287	99.89%
403	160	0.08%	200,447	99.97%
411	52	0.03%	200,499	100.00%
419	5	0.00%	200,504	100.00%

Table Q14. ELA Grade 4 Scale Score Frequency Distribution

Scale Score	N-Count	%	Cumulative	
			N-Count	%
138	2	0.00%	2	0.00%
146	9	0.00%	11	0.01%
154	28	0.01%	39	0.02%
162	64	0.03%	103	0.05%
170	133	0.06%	236	0.11%
178	239	0.12%	475	0.23%
186	472	0.23%	947	0.46%
194	778	0.38%	1,725	0.84%
202	1,029	0.50%	2,754	1.34%
210	1,313	0.64%	4,067	1.98%
218	1,564	0.76%	5,631	2.74%
226	1,881	0.92%	7,512	3.66%
232	2,043	1.00%	9,555	4.66%
237	2,435	1.19%	11,990	5.84%
242	2,602	1.27%	14,592	7.11%
247	2,805	1.37%	17,397	8.48%
251	3,018	1.47%	20,415	9.95%
255	3,421	1.67%	23,836	11.61%
259	3,825	1.86%	27,661	13.48%
263	3,983	1.94%	31,644	15.42%
266	4,335	2.11%	35,979	17.53%
270	4,679	2.28%	40,658	19.81%
273	5,035	2.45%	45,693	22.27%
276	5,392	2.63%	51,085	24.89%
280	5,674	2.76%	56,759	27.66%

Table Q14. ELA Grade 4 Scale Score Frequency Distribution (cont.)

Scale Score	N-Count	%	Cumulative	
			N-Count	%
283	6,211	3.03%	62,970	30.68%
287	6,515	3.17%	69,485	33.86%
289	6,714	3.27%	76,199	37.13%
292	7,035	3.43%	83,234	40.56%
296	7,513	3.66%	90,747	44.22%
299	7,700	3.75%	98,447	47.97%
302	7,904	3.85%	106,351	51.82%
305	8,017	3.91%	114,368	55.73%
308	8,113	3.95%	122,481	59.68%
311	8,074	3.93%	130,555	63.62%
315	7,997	3.90%	138,552	67.51%
320	7,756	3.78%	146,308	71.29%
321	7,537	3.67%	153,845	74.97%
325	7,294	3.55%	161,139	78.52%
328	6,829	3.33%	167,968	81.85%
332	6,313	3.08%	174,281	84.92%
335	5,691	2.77%	179,972	87.70%
339	5,148	2.51%	185,120	90.21%
343	4,429	2.16%	189,549	92.36%
347	3,895	1.90%	193,444	94.26%
351	3,242	1.58%	196,686	95.84%
356	2,636	1.28%	199,322	97.13%
361	2,000	0.97%	201,322	98.10%
367	1,507	0.73%	202,829	98.83%
374	1,077	0.52%	203,906	99.36%
381	694	0.34%	204,600	99.70%
391	362	0.18%	204,962	99.87%
399	176	0.09%	205,138	99.96%
407	63	0.03%	205,201	99.99%
415	15	0.01%	205,216	100.00%
423	5	0.00%	205,221	100.00%

Table Q15. ELA Grade 5 Scale Score Frequency Distribution

Scale Score	N-Count	%	Cumulative	
			N-Count	%
105	25	0.01%	25	0.01%
113	5	0.00%	30	0.01%
121	11	0.01%	41	0.02%
129	12	0.01%	53	0.03%
137	25	0.01%	78	0.04%
145	51	0.03%	129	0.06%
153	84	0.04%	213	0.11%
161	174	0.09%	387	0.19%
169	255	0.13%	642	0.32%
177	411	0.20%	1,053	0.52%
185	606	0.30%	1,659	0.83%
193	792	0.39%	2,451	1.22%
201	1,030	0.51%	3,481	1.73%
209	1,285	0.64%	4,766	2.37%
217	1,525	0.76%	6,291	3.13%
225	1,622	0.81%	7,913	3.94%
230	1,861	0.93%	9,774	4.86%
235	2,022	1.01%	11,796	5.87%
239	2,149	1.07%	13,945	6.94%
243	2,440	1.21%	16,385	8.16%
247	2,486	1.24%	18,871	9.39%
251	2,768	1.38%	21,639	10.77%
254	2,968	1.48%	24,607	12.25%
258	3,177	1.58%	27,784	13.83%
261	3,453	1.72%	31,237	15.55%
264	3,607	1.80%	34,844	17.34%
267	3,992	1.99%	38,836	19.33%
270	4,028	2.00%	42,864	21.34%
273	4,358	2.17%	47,222	23.50%
276	4,687	2.33%	51,909	25.84%
278	4,797	2.39%	56,706	28.22%
281	4,915	2.45%	61,621	30.67%
284	5,161	2.57%	66,782	33.24%
286	5,386	2.68%	72,168	35.92%
289	5,371	2.67%	77,539	38.59%
292	5,546	2.76%	83,085	41.35%
294	5,699	2.84%	88,784	44.19%
297	5,716	2.85%	94,500	47.04%
299	5,812	2.89%	100,312	49.93%
302	6,146	3.06%	106,458	52.99%

Table Q15. ELA Grade 5 Scale Score Frequency Distribution (cont.)

Scale Score	N-Count	%	Cumulative	
			N-Count	%
304	5,998	2.99%	112,456	55.97%
307	6,082	3.03%	118,538	59.00%
309	6,186	3.08%	124,724	62.08%
312	6,083	3.03%	130,807	65.11%
314	5,951	2.96%	136,758	68.07%
317	5,837	2.91%	142,595	70.97%
320	5,717	2.85%	148,312	73.82%
322	5,547	2.76%	153,859	76.58%
325	5,474	2.72%	159,333	79.31%
327	5,191	2.58%	164,524	81.89%
330	4,819	2.40%	169,343	84.29%
333	4,619	2.30%	173,962	86.59%
337	4,346	2.16%	178,308	88.75%
340	4,018	2.00%	182,326	90.75%
346	3,651	1.82%	185,977	92.57%
347	3,244	1.61%	189,221	94.18%
351	2,757	1.37%	191,978	95.55%
355	2,393	1.19%	194,371	96.75%
359	1,973	0.98%	196,344	97.73%
364	1,552	0.77%	197,896	98.50%
370	1,145	0.57%	199,041	99.07%
376	786	0.39%	199,827	99.46%
384	569	0.28%	200,396	99.74%
393	321	0.16%	200,717	99.90%
401	128	0.06%	200,845	99.97%
409	57	0.03%	200,902	100.00%
417	7	0.00%	200,909	100.00%

Table Q16. ELA Grade 6 Scale Score Frequency Distribution

Scale Score	N-Count	%	Cumulative	
			N-Count	%
118	5	0.00%	5	0.00%
126	14	0.01%	19	0.01%
134	11	0.01%	30	0.01%
142	24	0.01%	54	0.03%
150	32	0.02%	86	0.04%
158	72	0.04%	158	0.08%
166	124	0.06%	282	0.14%
174	230	0.11%	512	0.25%
183	356	0.18%	868	0.43%
191	536	0.27%	1,404	0.69%
199	785	0.39%	2,189	1.08%
207	888	0.44%	3,077	1.52%
215	1,208	0.60%	4,285	2.12%
223	1,369	0.68%	5,654	2.80%
228	1,481	0.73%	7,135	3.53%
233	1,722	0.85%	8,857	4.38%
237	1,855	0.92%	10,712	5.30%
241	2,072	1.03%	12,784	6.33%
245	2,140	1.06%	14,924	7.39%
248	2,353	1.16%	17,277	8.55%
252	2,530	1.25%	19,807	9.80%
255	2,757	1.36%	22,564	11.17%
258	2,890	1.43%	25,454	12.60%
261	2,988	1.48%	28,442	14.08%
264	3,182	1.58%	31,624	15.65%
266	3,455	1.71%	35,079	17.36%
269	3,644	1.80%	38,723	19.17%
272	3,871	1.92%	42,594	21.08%
274	3,861	1.91%	46,455	23.00%
277	4,119	2.04%	50,574	25.03%
279	4,312	2.13%	54,886	27.17%
283	4,453	2.20%	59,339	29.37%
284	4,675	2.31%	64,014	31.69%
286	4,684	2.32%	68,698	34.01%
288	4,877	2.41%	73,575	36.42%
290	5,106	2.53%	78,681	38.95%
293	5,112	2.53%	83,793	41.48%
295	5,199	2.57%	88,992	44.05%
297	5,440	2.69%	94,432	46.74%
299	5,464	2.70%	99,896	49.45%

Table Q16. ELA Grade 6 Scale Score Frequency Distribution (cont.)

Scale Score	N-Count	%	Cumulative	
			N-Count	%
301	5,636	2.79%	105,532	52.24%
304	5,587	2.77%	111,119	55.00%
306	5,681	2.81%	116,800	57.82%
308	5,740	2.84%	122,540	60.66%
310	5,702	2.82%	128,242	63.48%
313	5,866	2.90%	134,108	66.38%
315	5,774	2.86%	139,882	69.24%
317	5,562	2.75%	145,444	71.99%
320	5,509	2.73%	150,953	74.72%
322	5,366	2.66%	156,319	77.38%
325	5,216	2.58%	161,535	79.96%
327	4,972	2.46%	166,507	82.42%
330	4,806	2.38%	171,313	84.80%
333	4,667	2.31%	175,980	87.11%
338	4,348	2.15%	180,328	89.26%
340	4,003	1.98%	184,331	91.24%
343	3,563	1.76%	187,894	93.01%
347	3,219	1.59%	191,113	94.60%
351	2,790	1.38%	193,903	95.98%
356	2,360	1.17%	196,263	97.15%
361	1,978	0.98%	198,241	98.13%
367	1,427	0.71%	199,668	98.83%
374	1,038	0.51%	200,706	99.35%
383	724	0.36%	201,430	99.71%
396	399	0.20%	201,829	99.90%
404	156	0.08%	201,985	99.98%
412	37	0.02%	202,022	100.00%

Table Q17. ELA Grade 7 Scale Score Frequency Distribution

Scale Score	N-Count	%	Cumulative	
			N-Count	%
113	5	0.00%	5	0.00%
121	5	0.00%	10	0.01%
129	12	0.01%	22	0.01%
137	19	0.01%	41	0.02%
145	34	0.02%	75	0.04%
153	61	0.03%	136	0.07%
161	135	0.07%	271	0.14%
169	207	0.10%	478	0.24%
177	353	0.18%	831	0.42%
185	535	0.27%	1,366	0.68%
193	722	0.36%	2,088	1.04%
201	945	0.47%	3,033	1.52%
209	1,181	0.59%	4,214	2.11%
217	1,485	0.74%	5,699	2.85%
225	1,671	0.84%	7,370	3.69%
231	1,883	0.94%	9,253	4.63%
235	2,222	1.11%	11,475	5.74%
240	2,381	1.19%	13,856	6.93%
244	2,677	1.34%	16,533	8.27%
248	2,955	1.48%	19,488	9.75%
252	3,152	1.58%	22,640	11.33%
255	3,487	1.74%	26,127	13.07%
258	3,588	1.79%	29,715	14.86%
262	3,795	1.90%	33,510	16.76%
265	3,945	1.97%	37,455	18.74%
268	4,229	2.12%	41,684	20.85%
270	4,424	2.21%	46,108	23.06%
273	4,672	2.34%	50,780	25.40%
276	4,706	2.35%	55,486	27.76%
278	4,917	2.46%	60,403	30.21%
281	4,988	2.50%	65,391	32.71%
283	5,244	2.62%	70,635	35.33%
287	5,185	2.59%	75,820	37.93%
288	5,483	2.74%	81,303	40.67%
291	5,510	2.76%	86,813	43.43%
293	5,515	2.76%	92,328	46.18%
295	5,705	2.85%	98,033	49.04%
298	5,541	2.77%	103,574	51.81%
300	5,670	2.84%	109,244	54.65%
303	5,666	2.83%	114,910	57.48%

Table Q17. ELA Grade 7 Scale Score Frequency Distribution (cont.)

Scale Score	N-Count	%	Cumulative	
			N-Count	%
305	5,689	2.85%	120,599	60.33%
307	5,753	2.88%	126,352	63.20%
310	5,668	2.84%	132,020	66.04%
312	5,544	2.77%	137,564	68.81%
314	5,554	2.78%	143,118	71.59%
318	5,480	2.74%	148,598	74.33%
320	5,404	2.70%	154,002	77.04%
322	5,202	2.60%	159,204	79.64%
325	5,085	2.54%	164,289	82.18%
328	4,743	2.37%	169,032	84.55%
330	4,362	2.18%	173,394	86.74%
333	4,271	2.14%	177,665	88.87%
336	3,951	1.98%	181,616	90.85%
340	3,588	1.79%	185,204	92.64%
343	3,147	1.57%	188,351	94.22%
347	2,837	1.42%	191,188	95.64%
351	2,544	1.27%	193,732	96.91%
355	1,970	0.99%	195,702	97.89%
360	1,487	0.74%	197,189	98.64%
366	1,128	0.56%	198,317	99.20%
372	761	0.38%	199,078	99.58%
380	466	0.23%	199,544	99.82%
390	234	0.12%	199,778	99.93%
398	103	0.05%	199,881	99.98%
406	26	0.01%	199,907	100.00%
414	4	0.00%	199,911	100.00%

Table Q18. ELA Grade 8 Scale Score Frequency Distribution

Scale Score	N-Count	%	Cumulative	
			N-Count	%
97	3	0.00%	3	0.00%
105	8	0.00%	11	0.01%
113	8	0.00%	19	0.01%
120	21	0.01%	40	0.02%
128	29	0.01%	69	0.03%
136	33	0.02%	102	0.05%
144	70	0.03%	172	0.08%
152	106	0.05%	278	0.13%
160	175	0.08%	453	0.22%
168	295	0.14%	748	0.36%
176	372	0.18%	1,120	0.54%
184	518	0.25%	1,638	0.79%
192	618	0.30%	2,256	1.09%
200	819	0.40%	3,075	1.49%
207	945	0.46%	4,020	1.95%
215	1,096	0.53%	5,116	2.48%
223	1,231	0.60%	6,347	3.08%
228	1,408	0.68%	7,755	3.76%
233	1,495	0.73%	9,250	4.49%
237	1,687	0.82%	10,937	5.31%
241	1,857	0.90%	12,794	6.21%
245	2,044	0.99%	14,838	7.20%
248	2,307	1.12%	17,145	8.32%
252	2,548	1.24%	19,693	9.55%
255	2,796	1.36%	22,489	10.91%
258	2,959	1.44%	25,448	12.35%
261	3,102	1.50%	28,550	13.85%
264	3,417	1.66%	31,967	15.51%
267	3,601	1.75%	35,568	17.25%
269	3,911	1.90%	39,479	19.15%
272	3,898	1.89%	43,377	21.04%
275	4,169	2.02%	47,546	23.06%
277	4,460	2.16%	52,006	25.23%
280	4,602	2.23%	56,608	27.46%
284	4,798	2.33%	61,406	29.79%
285	4,973	2.41%	66,379	32.20%
287	5,167	2.51%	71,546	34.71%
289	5,347	2.59%	76,893	37.30%
292	5,359	2.60%	82,252	39.90%
294	5,576	2.70%	87,828	42.61%

Table Q18. ELA Grade 8 Scale Score Frequency Distribution (cont.)

Scale Score	N-Count	%	Cumulative	
			N-Count	%
296	5,688	2.76%	93,516	45.37%
299	5,651	2.74%	99,167	48.11%
301	5,861	2.84%	105,028	50.95%
303	5,838	2.83%	110,866	53.78%
306	5,971	2.90%	116,837	56.68%
308	5,943	2.88%	122,780	59.56%
311	6,073	2.95%	128,853	62.51%
313	5,954	2.89%	134,807	65.40%
316	5,945	2.88%	140,752	68.28%
318	5,992	2.91%	146,744	71.19%
321	5,851	2.84%	152,595	74.02%
324	5,819	2.82%	158,414	76.85%
326	5,807	2.82%	164,221	79.66%
329	5,430	2.63%	169,651	82.30%
333	5,270	2.56%	174,921	84.86%
336	5,029	2.44%	179,950	87.30%
339	4,710	2.28%	184,660	89.58%
343	4,436	2.15%	189,096	91.73%
347	3,962	1.92%	193,058	93.65%
352	3,593	1.74%	196,651	95.40%
357	2,847	1.38%	199,498	96.78%
363	2,477	1.20%	201,975	97.98%
371	1,827	0.89%	203,802	98.87%
380	1,200	0.58%	205,002	99.45%
393	755	0.37%	205,757	99.81%
401	290	0.14%	206,047	99.95%
409	93	0.05%	206,140	100.00%

Table Q19. Mathematics Grade 3 Scale Score Frequency Distribution

Scale Score	N-Count	%	Cumulative	
			N-Count	%
128	4	0.00%	4	0.00%
136	3	0.00%	7	0.00%
144	11	0.01%	18	0.01%
152	18	0.01%	36	0.02%
160	35	0.02%	71	0.04%
168	98	0.05%	169	0.08%
176	171	0.08%	340	0.17%
184	314	0.16%	654	0.32%
192	481	0.24%	1,135	0.56%
200	733	0.36%	1,868	0.93%
208	992	0.49%	2,860	1.42%
216	1,267	0.63%	4,127	2.05%
224	1,512	0.75%	5,639	2.80%
230	1,717	0.85%	7,356	3.65%
236	1,917	0.95%	9,273	4.61%
240	2,081	1.03%	11,354	5.64%
245	2,210	1.10%	13,564	6.74%
249	2,385	1.18%	15,949	7.92%
253	2,567	1.27%	18,516	9.20%
256	2,871	1.43%	21,387	10.62%
260	2,926	1.45%	24,313	12.07%
263	3,160	1.57%	27,473	13.64%
266	3,420	1.70%	30,893	15.34%
269	3,438	1.71%	34,331	17.05%
272	3,664	1.82%	37,995	18.87%
275	3,973	1.97%	41,968	20.84%
278	3,998	1.99%	45,966	22.83%
280	4,183	2.08%	50,149	24.91%
283	4,261	2.12%	54,410	27.02%
285	4,370	2.17%	58,780	29.19%
288	4,524	2.25%	63,304	31.44%
290	4,561	2.27%	67,865	33.70%
292	4,631	2.30%	72,496	36.00%
294	4,747	2.36%	77,243	38.36%
296	4,724	2.35%	81,967	40.71%
298	4,835	2.40%	86,802	43.11%
301	4,788	2.38%	91,590	45.49%
303	4,989	2.48%	96,579	47.96%
305	5,117	2.54%	101,696	50.51%
307	4,905	2.44%	106,601	52.94%

Table Q19. Mathematics Grade 3 Scale Score Frequency Distribution (cont.)

Scale Score	N-Count	%	Cumulative	
			N-Count	%
309	5,205	2.58%	111,806	55.53%
311	5,161	2.56%	116,967	58.09%
314	5,253	2.61%	122,220	60.70%
315	5,213	2.59%	127,433	63.29%
318	5,215	2.59%	132,648	65.88%
320	5,262	2.61%	137,910	68.49%
322	5,201	2.58%	143,111	71.07%
325	5,237	2.60%	148,348	73.67%
327	5,286	2.63%	153,634	76.30%
330	5,364	2.66%	158,998	78.96%
333	5,306	2.64%	164,304	81.60%
337	5,315	2.64%	169,619	84.24%
340	5,063	2.51%	174,682	86.75%
344	4,968	2.47%	179,650	89.22%
349	4,717	2.34%	184,367	91.56%
354	4,334	2.15%	188,701	93.71%
361	4,022	2.00%	192,723	95.71%
370	3,362	1.67%	196,085	97.38%
382	2,698	1.34%	198,783	98.72%
390	1,815	0.90%	200,598	99.62%
399	759	0.38%	201,357	100.00%

Table Q20. Mathematics Grade 4 Scale Score Frequency Distribution

Scale Score	N-Count	%	Cumulative	
			N-Count	%
134	4	0.00%	4	0.00%
143	2	0.00%	6	0.00%
151	18	0.01%	24	0.01%
159	28	0.01%	52	0.03%
167	87	0.04%	139	0.07%
175	164	0.08%	303	0.15%
184	357	0.17%	660	0.32%
192	645	0.31%	1,305	0.63%
200	994	0.48%	2,299	1.12%
208	1,399	0.68%	3,698	1.80%
216	1,706	0.83%	5,404	2.62%
224	2,019	0.98%	7,423	3.60%
232	2,259	1.10%	9,682	4.70%
238	2,410	1.17%	12,092	5.87%
243	2,550	1.24%	14,642	7.11%
247	2,783	1.35%	17,425	8.46%
251	2,645	1.28%	20,070	9.74%
255	2,703	1.31%	22,773	11.05%
258	2,710	1.32%	25,483	12.37%
261	2,873	1.39%	28,356	13.76%
264	2,806	1.36%	31,162	15.13%
266	2,893	1.40%	34,055	16.53%
269	2,956	1.43%	37,011	17.97%
271	3,010	1.46%	40,021	19.43%
273	3,069	1.49%	43,090	20.92%
275	3,068	1.49%	46,158	22.41%
278	3,157	1.53%	49,315	23.94%
280	3,140	1.52%	52,455	25.46%
281	3,056	1.48%	55,511	26.95%
283	3,305	1.60%	58,816	28.55%
285	3,242	1.57%	62,058	30.12%
287	3,328	1.62%	65,386	31.74%
289	3,342	1.62%	68,728	33.36%
291	3,284	1.59%	72,012	34.96%
292	3,442	1.67%	75,454	36.63%
294	3,429	1.66%	78,883	38.29%
296	3,521	1.71%	82,404	40.00%
297	3,590	1.74%	85,994	41.74%
299	3,683	1.79%	89,677	43.53%
301	3,773	1.83%	93,450	45.36%

Table Q20. Mathematics Grade 4 Scale Score Frequency Distribution (cont.)

Scale Score	N-Count	%	Cumulative	
			N-Count	%
302	3,817	1.85%	97,267	47.22%
304	3,902	1.89%	101,169	49.11%
306	3,963	1.92%	105,132	51.03%
307	3,903	1.89%	109,035	52.93%
309	4,001	1.94%	113,036	54.87%
311	4,068	1.97%	117,104	56.84%
313	4,156	2.02%	121,260	58.86%
314	4,160	2.02%	125,420	60.88%
316	4,296	2.09%	129,716	62.97%
318	4,233	2.05%	133,949	65.02%
320	4,299	2.09%	138,248	67.11%
322	4,462	2.17%	142,710	69.27%
324	4,429	2.15%	147,139	71.42%
327	4,593	2.23%	151,732	73.65%
329	4,605	2.24%	156,337	75.89%
332	4,647	2.26%	160,984	78.14%
334	4,767	2.31%	165,751	80.46%
337	4,760	2.31%	170,511	82.77%
341	4,813	2.34%	175,324	85.11%
344	4,731	2.30%	180,055	87.40%
348	4,670	2.27%	184,725	89.67%
353	4,550	2.21%	189,275	91.88%
359	4,468	2.17%	193,743	94.05%
366	4,110	2.00%	197,853	96.04%
377	3,648	1.77%	201,501	97.81%
396	2,855	1.39%	204,356	99.20%
404	1,651	0.80%	206,007	100.00%

Table Q21. Mathematics Grade 5 Scale Score Frequency Distribution

Scale Score	N-Count	%	Cumulative	
			N-Count	%
125	4	0.00%	4	0.00%
133	7	0.00%	11	0.01%
141	4	0.00%	15	0.01%
149	17	0.01%	32	0.02%
157	16	0.01%	48	0.02%
165	40	0.02%	88	0.04%
173	117	0.06%	205	0.10%
182	249	0.12%	454	0.23%
190	443	0.22%	897	0.45%
198	671	0.33%	1,568	0.78%
206	1,047	0.52%	2,615	1.30%
214	1,369	0.68%	3,984	1.99%
222	1,791	0.89%	5,775	2.88%
230	2,023	1.01%	7,798	3.89%
238	2,456	1.22%	10,254	5.11%
244	2,646	1.32%	12,900	6.43%
249	2,823	1.41%	15,723	7.84%
254	2,917	1.45%	18,640	9.29%
258	3,132	1.56%	21,772	10.85%
262	3,261	1.63%	25,033	12.48%
266	3,520	1.75%	28,553	14.24%
269	3,439	1.71%	31,992	15.95%
272	3,748	1.87%	35,740	17.82%
275	3,855	1.92%	39,595	19.74%
278	4,099	2.04%	43,694	21.78%
281	4,208	2.10%	47,902	23.88%
283	4,156	2.07%	52,058	25.95%
286	4,225	2.11%	56,283	28.06%
288	4,354	2.17%	60,637	30.23%
291	4,532	2.26%	65,169	32.49%
294	4,511	2.25%	69,680	34.74%
295	4,439	2.21%	74,119	36.95%
297	4,393	2.19%	78,512	39.14%
299	4,590	2.29%	83,102	41.43%
302	4,500	2.24%	87,602	43.68%
304	4,479	2.23%	92,081	45.91%
306	4,413	2.20%	96,494	48.11%
308	4,360	2.17%	100,854	50.28%
310	4,298	2.14%	105,152	52.43%
311	4,390	2.19%	109,542	54.61%

Table Q21. Mathematics Grade 5 Scale Score Frequency Distribution (cont.)

Scale Score	N-Count	%	Cumulative	
			N-Count	%
313	4,536	2.26%	114,078	56.88%
315	4,389	2.19%	118,467	59.06%
317	4,363	2.18%	122,830	61.24%
319	4,328	2.16%	127,158	63.40%
321	4,260	2.12%	131,418	65.52%
323	4,032	2.01%	135,450	67.53%
325	4,118	2.05%	139,568	69.58%
327	4,091	2.04%	143,659	71.62%
329	4,114	2.05%	147,773	73.68%
331	4,060	2.02%	151,833	75.70%
333	4,052	2.02%	155,885	77.72%
335	3,873	1.93%	159,758	79.65%
338	3,825	1.91%	163,583	81.56%
340	3,756	1.87%	167,339	83.43%
342	3,619	1.80%	170,958	85.23%
346	3,539	1.76%	174,497	87.00%
347	3,278	1.63%	177,775	88.63%
350	3,191	1.59%	180,966	90.22%
353	3,060	1.53%	184,026	91.75%
357	2,871	1.43%	186,897	93.18%
360	2,710	1.35%	189,607	94.53%
365	2,504	1.25%	192,111	95.78%
370	2,288	1.14%	194,399	96.92%
377	2,085	1.04%	196,484	97.96%
386	1,778	0.89%	198,262	98.85%
394	1,329	0.66%	199,591	99.51%
402	982	0.49%	200,573	100.00%

Table Q22. Mathematics Grade 6 Scale Score Frequency Distribution

Scale Score	N-Count	%	Cumulative	
			N-Count	%
121	4	0.00%	4	0.00%
129	4	0.00%	8	0.00%
137	7	0.00%	15	0.01%
145	20	0.01%	35	0.02%
153	25	0.01%	60	0.03%
161	75	0.04%	135	0.07%
169	129	0.06%	264	0.13%
178	239	0.12%	503	0.25%
186	395	0.20%	898	0.45%
194	611	0.30%	1,509	0.75%
202	977	0.49%	2,486	1.24%
210	1,277	0.64%	3,763	1.88%
218	1,662	0.83%	5,425	2.70%
226	1,909	0.95%	7,334	3.66%
234	2,232	1.11%	9,566	4.77%
240	2,351	1.17%	11,917	5.94%
245	2,604	1.30%	14,521	7.24%
249	2,648	1.32%	17,169	8.56%
253	2,685	1.34%	19,854	9.90%
256	2,834	1.41%	22,688	11.31%
260	2,782	1.39%	25,470	12.69%
263	3,082	1.54%	28,552	14.23%
265	3,017	1.50%	31,569	15.73%
268	3,089	1.54%	34,658	17.27%
270	3,174	1.58%	37,832	18.86%
273	3,306	1.65%	41,138	20.50%
275	3,389	1.69%	44,527	22.19%
277	3,440	1.71%	47,967	23.91%
279	3,522	1.76%	51,489	25.66%
282	3,554	1.77%	55,043	27.43%
284	3,616	1.80%	58,659	29.24%
286	3,740	1.86%	62,399	31.10%
288	3,644	1.82%	66,043	32.92%
290	3,784	1.89%	69,827	34.80%
292	3,901	1.94%	73,728	36.75%
293	3,932	1.96%	77,660	38.71%
295	3,931	1.96%	81,591	40.66%
297	4,011	2.00%	85,602	42.66%
299	4,042	2.01%	89,644	44.68%
301	4,089	2.04%	93,733	46.72%

Table Q22. Mathematics Grade 6 Scale Score Frequency Distribution (cont.)

Scale Score	N-Count	%	Cumulative	
			N-Count	%
303	4,164	2.08%	97,897	48.79%
305	4,149	2.07%	102,046	50.86%
306	4,111	2.05%	106,157	52.91%
308	4,159	2.07%	110,316	54.98%
310	4,187	2.09%	114,503	57.07%
312	4,267	2.13%	118,770	59.19%
314	4,016	2.00%	122,786	61.20%
316	4,128	2.06%	126,914	63.25%
318	4,185	2.09%	131,099	65.34%
320	4,085	2.04%	135,184	67.37%
322	4,005	2.00%	139,189	69.37%
324	3,998	1.99%	143,187	71.36%
326	3,936	1.96%	147,123	73.33%
328	3,881	1.93%	151,004	75.26%
330	3,899	1.94%	154,903	77.20%
332	3,680	1.83%	158,583	79.04%
334	3,711	1.85%	162,294	80.89%
336	3,544	1.77%	165,838	82.65%
340	3,425	1.71%	169,263	84.36%
341	3,428	1.71%	172,691	86.07%
343	3,226	1.61%	175,917	87.68%
346	3,123	1.56%	179,040	89.23%
349	2,981	1.49%	182,021	90.72%
352	2,773	1.38%	184,794	92.10%
355	2,720	1.36%	187,514	93.46%
358	2,542	1.27%	190,056	94.72%
363	2,257	1.12%	192,313	95.85%
367	2,077	1.04%	194,390	96.88%
373	1,888	0.94%	196,278	97.82%
381	1,602	0.80%	197,880	98.62%
392	1,317	0.66%	199,197	99.28%
400	902	0.45%	200,099	99.73%
408	546	0.27%	200,645	100.00%

Table Q23. Mathematics Grade 7 Scale Score Frequency Distribution

Scale Score	N-Count	%	Cumulative	
			N-Count	%
134	12	0.01%	12	0.01%
142	11	0.01%	23	0.01%
149	11	0.01%	34	0.02%
157	17	0.01%	51	0.03%
165	31	0.02%	82	0.04%
173	76	0.04%	158	0.08%
181	155	0.08%	313	0.16%
189	300	0.15%	613	0.31%
196	490	0.25%	1,103	0.56%
204	819	0.42%	1,922	0.98%
212	1,340	0.69%	3,262	1.67%
220	1,719	0.88%	4,981	2.55%
228	2,294	1.17%	7,275	3.72%
235	2,889	1.48%	10,164	5.20%
244	3,183	1.63%	13,347	6.83%
251	3,549	1.82%	16,896	8.65%
257	3,895	1.99%	20,791	10.64%
261	3,968	2.03%	24,759	12.67%
265	4,182	2.14%	28,941	14.81%
269	4,183	2.14%	33,124	16.95%
272	4,180	2.14%	37,304	19.09%
275	4,087	2.09%	41,391	21.18%
278	4,122	2.11%	45,513	23.29%
280	4,088	2.09%	49,601	25.39%
282	4,003	2.05%	53,604	27.44%
285	4,095	2.10%	57,699	29.53%
287	3,808	1.95%	61,507	31.48%
289	3,900	2.00%	65,407	33.48%
291	3,901	2.00%	69,308	35.47%
293	3,817	1.95%	73,125	37.43%
295	3,820	1.96%	76,945	39.38%
296	3,700	1.89%	80,645	41.28%
298	3,728	1.91%	84,373	43.18%
300	3,727	1.91%	88,100	45.09%
301	3,686	1.89%	91,786	46.98%
303	3,721	1.90%	95,507	48.88%
305	3,652	1.87%	99,159	50.75%
306	3,639	1.86%	102,798	52.61%
308	3,545	1.81%	106,343	54.43%
309	3,407	1.74%	109,750	56.17%

Table Q23. Mathematics Grade 7 Scale Score Frequency Distribution (cont.)

Scale Score	N-Count	%	Cumulative	
			N-Count	%
311	3,606	1.85%	113,356	58.02%
312	3,447	1.76%	116,803	59.78%
314	3,427	1.75%	120,230	61.54%
315	3,382	1.73%	123,612	63.27%
317	3,273	1.68%	126,885	64.94%
318	3,320	1.70%	130,205	66.64%
320	3,278	1.68%	133,483	68.32%
322	3,189	1.63%	136,672	69.95%
323	3,238	1.66%	139,910	71.61%
324	3,156	1.62%	143,066	73.22%
326	3,098	1.59%	146,164	74.81%
327	3,057	1.56%	149,221	76.37%
329	3,000	1.54%	152,221	77.91%
330	3,017	1.54%	155,238	79.45%
332	2,884	1.48%	158,122	80.93%
334	2,868	1.47%	160,990	82.40%
335	2,907	1.49%	163,897	83.89%
337	2,768	1.42%	166,665	85.30%
339	2,694	1.38%	169,359	86.68%
341	2,650	1.36%	172,009	88.04%
343	2,648	1.36%	174,657	89.39%
345	2,515	1.29%	177,172	90.68%
348	2,395	1.23%	179,567	91.91%
349	2,310	1.18%	181,877	93.09%
352	2,272	1.16%	184,149	94.25%
355	2,107	1.08%	186,256	95.33%
358	2,022	1.03%	188,278	96.36%
362	1,878	0.96%	190,156	97.33%
367	1,646	0.84%	191,802	98.17%
374	1,414	0.72%	193,216	98.89%
384	1,123	0.57%	194,339	99.47%
392	722	0.37%	195,061	99.84%
400	321	0.16%	195,382	100.00%

Table Q24. Mathematics Grade 8 Scale Score Frequency Distribution

Scale Score	N-Count	%	Cumulative	
			N-Count	%
117	10	0.01%	10	0.01%
125	15	0.01%	25	0.02%
133	11	0.01%	36	0.02%
141	16	0.01%	52	0.03%
149	18	0.01%	70	0.04%
157	57	0.04%	127	0.08%
164	82	0.05%	209	0.13%
172	167	0.10%	376	0.23%
180	316	0.20%	692	0.43%
188	511	0.32%	1,203	0.74%
196	822	0.51%	2,025	1.25%
204	1,198	0.74%	3,223	1.99%
212	1,559	0.96%	4,782	2.96%
220	2,022	1.25%	6,804	4.21%
228	2,334	1.44%	9,138	5.65%
236	2,748	1.70%	11,886	7.35%
243	3,128	1.93%	15,014	9.28%
248	3,254	2.01%	18,268	11.29%
253	3,521	2.18%	21,789	13.47%
257	3,548	2.19%	25,337	15.66%
261	3,651	2.26%	28,988	17.92%
265	3,771	2.33%	32,759	20.25%
268	3,784	2.34%	36,543	22.59%
271	3,916	2.42%	40,459	25.01%
274	3,785	2.34%	44,244	27.35%
276	3,798	2.35%	48,042	29.70%
279	3,814	2.36%	51,856	32.06%
281	3,909	2.42%	55,765	34.47%
283	3,891	2.41%	59,656	36.88%
287	3,818	2.36%	63,474	39.24%
288	3,727	2.30%	67,201	41.54%
290	3,627	2.24%	70,828	43.79%
292	3,697	2.29%	74,525	46.07%
293	3,639	2.25%	78,164	48.32%
295	3,602	2.23%	81,766	50.55%
297	3,544	2.19%	85,310	52.74%
299	3,460	2.14%	88,770	54.88%
301	3,354	2.07%	92,124	56.95%
302	3,356	2.07%	95,480	59.03%
304	3,313	2.05%	98,793	61.08%

Table Q24. Mathematics Grade 8 Scale Score Frequency Distribution (cont.)

Scale Score	N-Count	%	Cumulative	
			N-Count	%
306	3,307	2.04%	102,100	63.12%
307	3,332	2.06%	105,432	65.18%
309	3,230	2.00%	108,662	67.18%
311	3,091	1.91%	111,753	69.09%
313	3,059	1.89%	114,812	70.98%
314	3,014	1.86%	117,826	72.84%
316	2,958	1.83%	120,784	74.67%
318	2,744	1.70%	123,528	76.37%
319	2,729	1.69%	126,257	78.05%
322	2,796	1.73%	129,053	79.78%
323	2,753	1.70%	131,806	81.48%
325	2,429	1.50%	134,235	82.99%
326	2,503	1.55%	136,738	84.53%
328	2,422	1.50%	139,160	86.03%
330	2,262	1.40%	141,422	87.43%
332	2,172	1.34%	143,594	88.77%
334	1,986	1.23%	145,580	90.00%
336	1,937	1.20%	147,517	91.20%
339	1,837	1.14%	149,354	92.33%
341	1,595	0.99%	150,949	93.32%
344	1,490	0.92%	152,439	94.24%
346	1,473	0.91%	153,912	95.15%
349	1,302	0.80%	155,214	95.96%
353	1,188	0.73%	156,402	96.69%
357	1,116	0.69%	157,518	97.38%
361	984	0.61%	158,502	97.99%
366	840	0.52%	159,342	98.51%
372	714	0.44%	160,056	98.95%
380	608	0.38%	160,664	99.32%
391	482	0.30%	161,146	99.62%
399	350	0.22%	161,496	99.84%
407	187	0.12%	161,683	99.95%
415	73	0.05%	161,756	100.00%